

ARKISTOJEN KUVAILUMALLIT AINEISTON HAETTAVUUDEN NÄKÖKULMASTA

Piia Naukkarinen

Tampereen yliopisto
Informaatiotieteiden yksikkö
Informaatiotutkimus ja interaktiivinen
media
Pro gradu -tutkielma
Huhtikuu 2015

TAMPEREEN YLIOPISTO, Informaatitieteiden yksikkö

Informaatiotutkimus ja interaktiivinen media

NAUKKARINEN PIIA: Arkistojen kuvailumallit aineiston haettavuuden näkökulmasta

Pro gradu -tutkielma, 56 s.

Huhtikuu 2015

Tiedonhaku on muuttunut internetin ja sen hakukoneiden myötä. Tämä on vaikuttanut myös arkistoaineiston hakutapoihin. Arkistoaineiston käyttäjät olettavat verkossa olevien arkistoaineistohakukoneiden toimivan samaan tapaan kuin muutkin hakukoneet. Arkistoaineiston käyttäjiä kiinnostaa sisältö enemmän kuin se miten aineisto on järjestetty. Edellä mainittujen vuoksi aineiston intellektuaalisella kuvailulla on merkittävä rooli aineiston haun onnistumisessa. Arkistoaineistojen käyttäjillä on tutkimusten mukaan useita erilaisia hakukriteereitä, joilla he etsivät itselleen relevantteja aineistoja. Hakukriteerit vaihtelevat jonkin verran käyttäjäryhmän mukaan.

Arkistokuvailu on muutoksessa. Kehitteillä on uusi käsitteellinen arkistokuvailumalli, jossa suurimpana muutoksena aikaisempaan on elementtien kuvailu erillisinä ja niiden yhdistäminen suhteiden avulla. Tämä ja muu linkitetty data helpottaa tutkimusten mukaan mm. aiheen mukaista pääsyä aineistoihin (Lindquist, Dulock, Törnroos, Hyvönen, & Mäkelä, 2013).

Tässä tutkielmassa verrataan aiemmista tutkimuksista löydettyjä arkistoaineiston käyttäjien hakukriteereitä nykyiseen arkistokuvailumalliin ja JHS 143 -suosituksen sekä SÄHKE2-määräyksen metatietomalleihin. Hakukriteerejä verrataan myös luonnosvaiheessa olevaan uuteen AHAA-kuvailumalliin sekä siitä johdettuihin keskeneräisiin kuvailusääntöihin mahdollisuuksien puitteissa.

Tutkielman tulosten perusteella aineiston haettavuus ja löydettävyyys paranevat uudenlaisen kuvailun myötä. Aineiston entiteettien kuvailu erillisinä ja niiden linkittäminen toisiinsa ja muihin liittyviin aineistoihin kasvattaa aineiston löytymisen mahdollisuuksia. Aineiston kuvailuentiteettien tallentaminen auktorisoituna tietona ontologiaan lisää kuvailun johdonmukaisuutta ja tarkuutta.

SISÄLLYS

1.	JOHDANTO.....	1
2.	ARKISTOKUVAILU	3
2.1.	Kuvailun peruskäsitteitä ja lähtökohtia.....	3
2.2.	Arkistokuvailustandardit ja kuvailumetatieto	4
2.2.1.	Julkishallinnon asiakirjojen kuvailun ja hallinnan metatiedot (JHS 143)	7
2.2.2.	SÄHKE2 ja sen metatietomalli.....	11
2.3.	Bibliografinen kuvailu, FRBR käsitteellisenä mallina ja RDA-säännöstö	13
2.4.	Kulttuuriperintöaineistojen kuvailun yhdistäminen ja AHAA-käsittemalli.....	16
3.	KATSAUS TIEDONHAKUTUTKIMUKSEEN JA -KIRJALLISUUTEEN	22
3.1.	Tiedonhankinta ja tiedonhaku.....	22
3.2.	Arkistonkäyttäjien informaatiotarpeet ja hakukäyttäytyminen	24
3.3.	Linkitetty data ja ontologiat haun tukena.....	29
4.	TUTKIMUSASETELMA	30
4.1.	Tutkimuskysymykset.....	30
4.2.	Hakukriteerien kerääminen ja tutkielmaan valitut kuvailumallit	31
4.3.	Tutkielma-aineiston analyysi	31
5.	TULOKSET.....	34
5.1.	Arkistoaineiston käyttäjien hakukriteerit	34
5.2.	Nykyisen kuvailumallin, JHS 143 -metatietomallin ja SÄHKE2-metatietomallin vertailu hakukriteereihin..	37
5.3.	AHAA-kuvailumallin ja -kuvailusääntöjen vertailu hakukriteereihin	40
5.4.	Kuvailumallien vertailua.....	45
6.	TULOSTEN TARKASTELUA	47
7.	JOHTOPÄÄTÖKSET	50
7.1.	Kuvailumallien vaikutuksia aineiston haettavuuteen.....	50
7.2.	Tulosten pohjalta annettavat suositukset.....	51
7.3.	Tuloksiin ja menetelmiin liittyvät rajoitukset	52
	LÄHTEET	53

1. JOHDANTO

Arkistoaineistojen käyttäjiä on monenlaisia ja heidän hakutarpeensa eroavat jossain määrin toisistaan. Käyttäjien arkistoaineistojen hakua voi parantaa monien tutkijoiden mukaan parantamalla ja lisäämällä aineistosta tallennettavaa kuvailumetatietoa (Schaffner, 2009, 4). Tämän perusteella aineistojen löydettävyyteen voidaan merkittävästi vaikuttaa parantamalla kuvailumallia ja kuvailustandardeja. Arkistoaineistojen ja muiden kulttuuriperintöaineistojen käyttäjät haluavat etsiä nykyään tietoa useimmiten verkossa olevien hakupalveluiden kautta ja käyttäjät olettavat aineistojen olevan käytössä heille sopivalla tavalla (Adams, 2007, 31; Huvila, 2008, 17).

Arkistokuvailua uudistetaan parhaillaan käsitteellisellä tasolla ja siitä johdetaan myös uuden mallin mukaisia kuvailusääntöjä, joiden mukaisesti aineistojen kuvailu toteutetaan. Uudistamistyö tähtää muun muassa aineistojen parempaan löydettävyyteen ja aineistojen suurempaan käyttöön. (Henttonen & Kilkki, 2013, 4.) Kirjastojen, arkistojen ja museoiden tietojärjestelmiä on Suomessakin pyritty yhdistämään jo 1990-luvulta lähtien mm. Kamut-projektissa, ja tässä yhteydessä on todettu yhtenäisten tietorakenteiden tarve (KAMUT, 1997, 2). Kirjastokuvailussa käsittemalliksi hyväksytty FRBR-malli sekä siitä johdettu RDA-säännöstö ovat olleet pohjana myös arkistokuvailua uudistettaessa (Henttonen & Kilkki, 2013, 18). AHAA-hankkeessa arkistokuvailua pyritään kehittämään vastaamaan uuden toimintaympäristön vaatimuksia (Henttonen & Kilkki, 2013, 2).

Tässä tutkielmassa selvitetään millaisia hakutarpeita ja tiedontarpeita arkistoaineistojen käyttäjillä aikaisempien tutkimusten mukaan on. Tutkielmassa käydään läpi uusimpia tutkimustuloksia arkistoaineistojen käyttäjien tiedonhakutarpeista. Tutkimuksista ja kirjallisuudesta löydettyjä arkistoaineistojen käyttäjien havaittuja hakukriteereitä verrataan arkistokuvailun nykyiseen kuvailumalliin ja käytössä oleviin metatietomalleihin sekä luonnosvaiheessa olevaan uuteen AHAA-kuvailumalliin ja keskeneräisiin AHAA-kuvailusääntöihin.

Tutkielma jakautuu eri lukuihin seuraavasti. Luvussa 2 kerrotaan arkistojen ja kirjastojen kuvailumalleista sekä -standardeista. Luvussa 3 käydään läpi tiedonhaketutkimusten tuloksia ja arkistokäyttäjien tiedonhakutarpeita. Luvussa neljä kerrotaan tutkimusasetelma, tutkimuksen rajaukset sekä tutkimustavat. Luvussa 5 esitellään tiedonhaku- ja tiedontarvetutkimusten perusteella laadittu hakukriteerilista. Hakukriteerilistan perusteella laadittiin taulukko, jota verrataan kuvailumallien ja metatietomallien määrittämiin metatietoihin. Tämän jälkeen luvussa 6 kerrotaan tutkimuksen tulokset. Tulosten perusteella tehdyt johtopäätökset ja suositukset käydään läpi luvussa 7.

2. ARKISTOKUVAILU

Tässä luvussa kerrotaan arkistokuvailusta ja siihen vaikuttaneista tekijöistä. Luvussa kerrotaan myös kirjastokuvailusta, koska FRBR-käsitelmä on ollut pohjana RDA-kuvailusäännöstölle. RDA on vaikuttanut AHAA-kuvailumallin muodostumiseen KDK-hankkeen suosituksesta. Alaluvussa 2.2 kerrotaan suomalaisista arkistokuvailustandardeista sekä joistakin eroista kansainvälisiin kuvailustandardeihin. Alaluvussa 2.3 kerrotaan julkishallinnolle tehdystä JHS 143 -suosituksen asiakirjojen kuvailun ja hallinnan metatiedoista. Alaluvussa 2.4 on tietoa sähköisen asiakirjahallinnon SÄHKE2-määräyksen metatiedoista.

2.1. Kuvailun peruskäsitteitä ja lähtökohtia

Kulttuuriperintöaineistojen kuvailulle määritellään yhdenmukaisuuden varmistamiseksi aina ohjeita, jotka sisältävät tietoa muun muassa siitä, mitä tietoelementtejä kuvataan, miten tietoelementit koodataan erottumaan toisistaan ja mitkä tietoelementit ovat ensisijaisia hakuelementtejä (Alaterä & Halttunen, 2002, 49). Aineistojen kuvailulla on monia eri tarkoituksia (Alaterä & Halttunen, 2002, 47-48), tässä tutkielmassa käsitellään kuvailua aineiston haun ja löytämisen näkökulmasta. Aineistokuvailu jaetaan usein luettelointitietoihin ja sisällönkuvailuun. Sisällönkuvailun tarkoitus on yleensä kertoa suppeasti dokumentin tai aineiston aihepiiristä ja asiasisällöstä. Kuvailutiedot tallennetaan nykyisin haun mahdollistamiseksi sähköiseen tietokantaan. (Alaterä & Halttunen, 2002, 50 & 68.)

Arkistoaineisto on toimijan toiminnan dokumentaatiota. Arkistokuvailun tarkoituksena on tietovarannon löytäminen, tunnistaminen, valinta sekä käyttöön saaminen ja ymmärrettävyys. Näiden toteutumiseksi vaaditaan *metatietojen* kokonaisuus, joiden avulla myös asioiden väliset suhteet pystytään kuvaamaan. Arkistokuvailussa kuvailutietojen kokonaisuus muodostuu sekä tietovarantoa itseään että sen konteksteja kuvaavista metatiedoista. Kuvailun avulla dokumentoidaan aineiston *provenienssi* ja se on sekä fyysisen että elektronisen aineiston hallinnan väline. (Gracy, 2014, 7.) Metatieto-käsitettä käytetään tässä tutkielmassa yleisterminä kaikelle kulttuuriperintöaineiston kuvailutiedolle, kuten Alaterä ja Halttunen (2002, 48-49) sen määrittelevät. Provenienssi tarkoittaa organisaatiota tai henkilöä, joka on osallistunut arkistoaineiston luomiseen, keräämiseen tai käyttämiseen (Arkistolaitos, 2013). Arkistoaineistossa olennaisia ovat sen osien sisäiset suhdeverkostot. *Sarjaa* (aineiston osa) tai yksittäistä asiakirjaa ei kuvailla yksittäin ja niiden merkitys muodostuu suhteessa arkiston muihin sarjoihin ja asiakirjoihin. (Arkistolaitos, 2006.)

2.2. Arkistokuvailustandardit ja kuvailumetatieto

Arkistokuvailussa olennaista on aineiston monikerroksisuus. Kuvailun tarkoituksena on arkistoaineiston intellektuaalisen eli käsitteellisen hallinnan helpottaminen. Luettelointi sen sijaan tähtää aineiston fyysiseen hallintaan. Myös arkistokuvailussa pyritään yhdenmukaisiin käytäntöihin erilaisten standardien avulla aineiston haun ja käytettävyyden helpottamiseksi. Niitä ovat esimerkiksi Suomen arkistolaitoksen vuonna 1997 julkaisema Arkistojen kuvailu- ja luettelointisäännöt, joka on pääosin yhdenmukainen Kansainvälisen arkistoneuvoston (ICA, International Council on Archives) kuvailustandardien (ISAD(G) eli General International Standard Archival Description ja ISAAF(CPF) eli International Standard Archival Authority Record for Corporate Bodies, Persons and Families (joskus lyhennetty myös ISAAR(CPF)) kanssa. Arkistokuvailu tarkoittaa ”*tietoja, jotka identifioivat ja kuvaavat arkistonmuodostajan ja sen toiminnasta kertyneen aineiston*”. Kuvailu edesauttaa lisäksi aineiston ymmärtämistä, koska siinä kuvaillaan myös kontekstia, jossa tiedot ovat syntyneet. (Arkistolaitos, 2006.) Monitasoinen kuvailu ja arkistoluettelo muodostavat arkistohakemiston, joka mahdollistaa aineiston haun (Arkistolaitos, 1997, 9).

Arkistokuvailu koostuu yleensä neljästä hierarkkisesta tasosta, joista ylin on arkistonmuodostajataso. Sen alla ovat arkisto-, sarja- ja yksikkötasot. Kuvailutiedot jaetaan myös kuvailualueisiin ja *kuvailuelementteihin* (lueteltu alla). Kuvailualueiden- ja elementtien (ks. taulukko 1) tietoja merkitään eri kuvailutasoille, osa elementeistä sopii usealle eri tasolle. Myös arkistonmuodostajan yläpuolella oleva yleistaso on mahdollista kuvailla. (Arkistolaitos, 2006.) Vuonna 1997 julkaistu standardi on tarkoitettu vain tekstiaineistojen kuvailun ja luetteloinnin tueksi (Arkistolaitos, 1997, 8). Eri tasojen kuvailussa pyritään välttämään toistoa. Yhdellä tasolla kuvaillaan ne asiat, jotka koskevat kaikkia kyseisen tason asiakirjoja. Kuvailutasot liitetään toisiinsa hierarkkisesti. (Henttonen, 2012, 1-2.)

”Kuvailualueiden ja -elementtien mukaisia tietoja annetaan soveltuvien osien joko arkistonmuodostajasta tai aineistosta tai molemmista. Osa tiedoista on mielekkäitä vain tietyllä tasolla, esim. historia tai elämäkerta voi liittyä vain arkistonmuodostajaan.[...] Minimikuvailu koostuu yllä (tässä taulukko 1) olevassa ryhmittelyssä tähdellä merkityistä elementeistä. Oppikirjan lopussa on esimerkkejä kuvailuista ja niihin liittyvistä arkistoluetteloista. Arkistoyksikkö määritellään sarjaa pienemmäksi sisällönkuvailukokonaisuudeksi, joka rajataan kuvailun laatimisen yhteydessä.” (Arkistolaitos, 1997a)

Taulukko 1: Kuvailualueet ja -elementit (Arkistolaitos, 1997a, 15.)

Kuvailualue	Kuvailuelementit
1. Identifiointitietojen alue	
	1.1* Nimet
	1.2 Tunnus
	1.3* Aikamääreet
	1.4 Maantieteellinen toiminta-alue
2. Sisältötietojen alue	
	2.1* Historia tai elämäkerta
	2.2 Tehtävät
	2.3* Tietosisältö
	2.4 Arkistonmuodostusprosessi
	2.5* Hakusanat
3. Käytettävyystietojen alue	
	3.1 Järjestämisperiaatteet
	3.2 Hakemistot
	3.3* Aineiston laajuus
	3.4 Kunto ja fyysinen muoto
	3.5* Asiakirjakieli
	3.6* Käyttörajoitukset
	3.7 Käyttökopiot
4. Omistus- ja käsittelytietojen alue	
	4.1 Omistushistoria
	4.2 Luovutustiedot
	4.3 Järjestelytyöt
	4.4 Hävinnyt/hävitetty
	4.5 Kuvailutiedot
5. Viitetietojen alue	
	5.1 Julkaisut, tutkimukset
	5.2* Edeltäjät/seuraajat
	5.3 Liittyvät aineistot
	5.4 Muualla säilytettävä aineisto
	5.5 Huomautustiedot

Arkistokuvailun lähtökohtana ovat arkistonmuodostaja ja arkiston rakenneosat. Arkistoaineiston rakenneosat ovat ensisijaisesti loogisia kokonaisuuksia, eivät fyysisiä. Muodoltaan, tehtävältään tai sisällöltään yhteenkuuluvat asiakirjat muodostavat sarjan eli arkiston osan. Kuvailu saattaa olla ainoa mahdollisuus osoittaa asiakirjojen proveniensi esimerkiksi elektronisten aineistojen kohdalla. (Arkistolaitos, 1997a, 11.) Asiakirjat voivat liittyä useaan eri käyttökontekstiin (ja syntykontekstiin), joten niiden liittäminen vain yhteen *fondiin* (asiakirjojen kokonaisuus) ei luo riittävää kuvailutietoa. Riittävään arkistokuvailuun tarvitaan siis monimutkaisempi malli. (Henttonen, 2012, 4.) Nykyiset kuvailusäännöt ovat riittämättömiä ja niiden perusteet, tavoitteet ja sisällöt täytyy muokata uudelleen. Kuvailun uudistamista tehdään Arkistojen hakemistopalvelu-hankkeessa (AHAA) arkistolaitoksen johdolla. Yksi hankkeen tavoitteista on tuottaa uusi *käsitelmä* sekä siihen perustuvat uudet kuvailu- ja luettelointisäännöt. (Henttonen, 2012, 6.) Kuvailun käsitelmällä tarkoitetaan tässä tutkielmassa ylimmällä tasolla määriteltyjä kuvailun lähtökohtia. AHAA-kuvailumalli on luonteeltaan käsitelmä, toisin kuin nykyinen kuvailumalli vuodelta 1997. Nykyinen kuvailumalli ohjaa kuvailua konkreettisemmalla tasolla. Käsitelmän avulla voidaan luoda lähempänä käytännön toteutusta oleva *metatietomalli*, jota käytetään kuvailussa ja arkistohakemistojärjestelmissä. Metatietomallissa aineistosta tallennettavat tiedot kuvaillaan tarkemmin ja se liittyy kiinteästi kuvailusääntöihin.

Suomalaiset kuvailusäännöt eroavat kansainvälisistä kuvailustandardeista (ISAD(G), ISAAR(CPF), ISDF (International Standard for Describing Functions)) asiakirjoihin liittyvien *toimijoiden* kuvailussa. Kansainvälisten standardien puitteissa toimijoita on mahdollista kuvailla monipuolisemmin ja ne voivat olla liittyä asiakirjoihin muussakin kuin arkistonmuodostaja-suhteessa. Toimija eli organisaatio tai henkilö voi olla kuvailtuna esimerkiksi asiakirjan laatijana, sen aiheena tai esimerkiksi tekijänoikeuksien haltijana. (Henttonen, 2012, 3.) Organisaatioiden tehtävät ovat yleensä pitkäikäisempiä kuin itse organisaatiot. Tästä syystä ICA:n (International Council of Archives) standardien mukaan arkistot, niihin liittyvät toimijat ja niiden tehtävät kuvaillaan toisistaan erillään. Tehtävistä voi syntyä asiakirjoja moniin arkistoihin. Kuvailtavat entiteetit liitetään toisiinsa erilaisilla suhteilla. (Henttonen, 2012, 3-4.)

Kansainvälisten kuvailustandardien pysyvänä tavoitteena on johdonmukaiset, relevantit ja täsmälliset kuvailut. Kuvailujen tulee helpottaa arkistoaineistoinformaation hakua ja välitystä sekä mahdollistaa erillisten arkistojen kuvailutietojen tallentaminen yhdistettyyn informaatiojärjestelmään. ICA on kehittänyt neljä standardia: General International Standard Archival Description (ISAD(G)), International Standard Archival Authority Records–Corporate Bodies, Persons, and Families (ISAAR(CPF)), International Standard Description of Functions

(ISDF) ja International Standard Description of Institutions with Archival Holdings (ISDIAH). ICA:n tavoitteena on tällä hetkellä näiden standardien yhdistäminen kokonaisvaltaiseksi kuvailustandardiksi käsitteellisin menetelmin. Näistä vain ISAD(G) on otettu laajasti käyttöön arkistokuvailujärjestelmissä kansainvälisesti. (Gueguen, Manoel, Pitti, & Grimoüard, 2013, 568.)

Myös tekniset ratkaisut vaikuttavat aineiston löydettävyyteen. Suomalaisessa Vakka-tietokannassa käytetään relaatiotietokantaa kuvailutietojen tallentamiseen. Muissa maissa teknisten ratkaisujen pohjana on usein EAD (Encoded Archival Description), joka on yksi tapa esittää monitasoisen kuvailun tiedot XML-formaatissa (Henttonen, 2012, 4). Henttonen näkemyksen mukaan suomalaiset kuvailusäännöt ovat liian yksinkertaisia, mikäli ICA:n kuvailusääntöjen suunta on tavoitteena. Arkistokuvailun monimutkaisuus pitäisi kuitenkin piilottaa käyttäjiltä, jotta hakujärjestelmät olisivat helppokäyttöisempiä eikä niiden käyttämiseen tarvitsisi arkistojen tuntemusta. (Henttonen, 2012, 4-5.) EAD-formaatin myötä ajatus siitä, että aineistoja voisi hakea vain kokoelma-tasoisten katalogien kautta, muuttui. Tämä ja EAD:n käyttäminen edisti myös aineistojen haettavuutta yleisesti käytettyjen hakukoneiden kautta. (Gracy, 2014, 8.)

Digitaalisten kokoelmien kuvailussa voidaan käyttää sekä kontrolloitua sanastoa että luonnollisen kielen kuvailusanoja (Parandjuk, 2010, 129). Tähän sisältyy digitaalisten kokoelmien rakentaminen siten, että niitä voi etsiä aiheen, yleisön tai mediatyyppin mukaisesti (Morville & Rosenfeld Parandjukin, 2010, 129 mukaan). Näin ollen myös teknisten määrittelyiden ja metadatastandardien kehittäminen vaikuttavat palvelun informaatioarkkitehtuurin toimivuuteen digitaalisissa kirjastoissa, sillä niiden avulla kirjastojärjestelmien yhteentoimivuus paranee (Parandjuk, 2010, 125).

2.2.1. Julkishallinnon asiakirjojen kuvailun ja hallinnan metatiedot (JHS 143)

JHS 143 -suositus on tarkoitettu julkisen hallinnon asiakirjojen hallinnan ja säilyttämisen ohjeeksi. Asiakirjahallintajärjestelmiä suunniteltaessa tulisi ottaa huomioon suosituksessa määritellyt metatiedot ja muut ohjeet. (JHS 143, 2006, 2.) Suosituksessa kehoitetaan sisällönkuvailuun kohdistuvien metatietojen yhdenmukaisuuteen muun aineiston kanssa käyttämällä sanastoja sekä luokittelu- ja merkintäjärjestelmiä. Konteksti-kategorian metatietoja tarvitaan suosituksen mukaan tiedon ymmärrettävyyden ja tulkittavuuden säilyttämiseksi. (JHS 143, 2006, 4-5.) Taulukossa 2 on lueteltu kategorioittain asiakirjallisen tiedon JHS 143 -suosituksen metatietoelementit. Pakollisuus on suosituksessa jaettu kahteen osaan, metatieto voi olla pakollinen asiakirjahallinnassa, mutta julkaistavalle asiakirjalle se ei välttämättä ole pakollista ja toisinpäin

(JHS 143, 2006). Pakollisuus on merkitty taulukossa 2 merkillä x ja ehdollinen pakollisuus merkillä (x).

Taulukko 2: JHS 143 -suosituksen asiakirjallisen tiedon metatiedot.

Kategoria	Elementti	Pakollisuus	
		Asiakirjahallinto	Julkaiseminen
Paikallistaminen	Identifiointitunnus	x	x
	Sijaintipaikka	x	
Sisällönkuvailu	Nimeke	x	x
	Aihe		x
	Kuvaus		
	Kieli		
	Kohdeyleisö		
	Kattavuus		
	Lähde		
	Laji		
Käyttöedellytykset	Oikeudet	x	
	Julkisuus	x	
	Säilytysaika	x	
	Formaatti	x	
	Suojeluluokka		
Konteksti	Toimija	x	x
	Tehtävä	x	
	Asiakirjan tyyppi	(x)	
	Suhde	(x)	
	Valtuutus		
Elinkaari	Aikamääre	x	x
	Tila	x	
	Käsittelyhistoria	x	
	Säilytyshistoria	(x)	

Taulukossa 2 mainittava metatieto *asiakirjatyypin* on pakollinen, jos sitä käytetään asiakirjan säilytysajan määräytymisen perusteena. *Suhde* on pakollinen, jos suhde on olemassa ja sillä on merkitystä asiakirjan elinkaaren hallinnan kannalta, määritelmällisesti suhde kuvataan samaan aktiin liittyvien asiakirjojen välillä. *Säilytyshistoria* on pakollinen, jos asiakirjoja säilytetään sähköisessä muodossa niin pitkään, että niille joudutaan suorittamaan migraatiotoimenpiteitä. (JHS 143, 2006, 13-39.) Asiakirjojen metatietoelementeillä on myös tarkenteita, ne on lueteltu taulukossa 3.

Taulukko 3: JHS 143 Asiakirjojen kuvailun ja hallinnan metatietoelementit ja niiden tarkenteet

Elementit	Elementtien tarkenteet
Nimeke	
	Vaihtoehtoinen nimeke
Toimija	
	Rooli tekijä
	Yhteisön nimi
	Henkilön nimi
Aihe	
Kuvaus	
	Tiivistelmä
	Sisällysluettelo
Toimija	
	Rooli julkaisija
Toimija	
	Rooli muu tekijä
	Yhteisön nimi
	Henkilön nimi
Aikamääre	
	Laatimisaika
	Julkistamisaika
	Hyväksymisaika
	Saatavillaoloaika
	Vastaanottoaika
	Kokoamisaika
	Voimassaoloaika
	Muokkausaika
	Lähtämisaika
	Tekijänoikeusaika
	Rekisteröintiaika
	Arkistointiaika
Laji	
Formaatti	
	Koko
	Tietoväline
	Tiedostomuoto
	Salaustapa
	Pakkaustapa
	Laitteisto
Identifiointitunnus	
	Viite
Lähde	
Kieli	
Suhde	
	Uudempi versio
	Vanhempi versio
	Korvattu
	Korvaa

	Pakollinen
	Liittyy pakollisena
	Sisältää
	Kuuluu
	Viittaa
	On viitattu
	Formaatti (aikaisempi)
	Formaatti (myöhempi)
	Julkinen versio
	Salassa pidettävä versio
	Noudattaa
Kattavuus	
	Hallinnonala
	Alueellinen kattavuus
	Ajallinen kattavuus
Oikeudet	
	Oikeuden tyyppi
	Selite
Asiakirjan tyyppi	
Julkisuus	
	Julkisuusaste
	Turvaluokka
	Omistaja
	Henkilötiedot
	Käsittelyoikeuksien kuvaus
	Henkilö, jolla on käsittelyoikeudet
	Rooli, jolla on käsittelyoikeudet
	Salassapitoperuste
	Salassapitoaika
	Salassapidon päättymisajankohta
Toimija	
	Roolina jakelija
Kohdeyleisö	
	Vaatimustaso
	Välittäjä
Sijaintipaikka	
	Tilapäinen sijoitus
	Selite
Säilytysaika	
	Säilytysajan pituus
	Säilytysajan päättymisajankohta
	Säilytysajan peruste
	Hävitysajankohta
	Hävitystapa
	Hävittämisen peruste
	Hävityksen suorittaja
	Siirtoajankohta
Suojeluluokka	
Toimija	

	Yhteisön nimi
	Henkilön nimi
	Yhteisön tunniste
	Henkilön tunniste
	Toimijan rooli
	Toimivalta
Tehtävä	
	Tehtäväryhmätunnus
	Rekisteritunnus
Valtuutus	
	Säännös
	Kohde
	Selite
Tila	
	Valmistumisen tila
	Elinkaaren tila
	Versio
Käsittelyhistoria	
	Toimenpiteen ajankohta
	Toimenpiteen suorittaja
	Toimenpiteen tyyppi
	Selite
Säilytyshistoria	
	(SÄHKE-määräyksestä)

Metatietoelementtien tarkenteet kertovat, millaisia tietoja kyseiseen elementtiin on mahdollista tallentaa. JSH 143 -suositukset on laadittu ottaen huomioon sekä Dublin Core -metadataformaatti että arkistolaitoksen SÄHKE-metatietomäärittely (*JHS 143*, 2006, 6).

2.2.2.SÄHKE2 ja sen metatietomalli

Sähköisenä tallennettaville ja säilytettävälle julkishallinnon asiakirjallisille tiedoille on oma määräys nimeltään *SÄHKE* eli sähköisten asiakirjallisten tietojen käsittely, hallinta ja säilyttäminen. Sen uusin versio on SÄHKE2, jonka metatietomallista voi nähdä muun muassa mitä tietoja sähköisistä asiakirjoista tulee tallentaa.

Arkistolaitoksen laatimat SÄHKE-määräykset on tehty ohjaamaan julkishallinnon asiakirjahallintaa asiakirjahallintajärjestelmissä. Määräysten avulla organisaatiot voivat laatia elektronisten aineistojen arkistonmuodostussuunnitelman (eAMS) tiedon luotettavan sähköisen säilyttämisen varmistamiseksi. SÄHKE-määräykset määrittelevät vaatimuksia ja ominaisuuksia, joita tarvitaan tietojärjestelmiin sisältyvien tietojen säilyttämiselle pysyvästi sähköisessä muodossa. Määräykset koskevat valtionhallinnon ja kuntasektorin organisaatioita. (Arkistolaitos,

2013b.) SÄHKE-määräyksen laatimisessa on otettu huomioon SFS-ISO-standardeja ja MoReq2:n vaatimuksia sekä JHS 143 -suositukset (Arkistolaitos, 2009b, 6).

SÄHKE2:ssa määritellään pakolliset metatiedot sähköisenä säilytettävälle asiakirjoille (Arkistolaitos, 2009, 6-29.):

- Tehtävä (eAMS:in tehtävuokituksen mukainen)
- Toimenpiteen tyyppi (määräytyy eAMS:iin kuvattujen tehtävien käsittelyvaiheiden mukaisesti)
- Asiakirjan tyyppi (käyttötarkoitus)
- Tila (käsittelyprosessin tila)
- Henkilötietoja (valinta seuraavista: ei sisällä henkilötietoja, sisältää henkilötietoja, sisältää arkaluontoisia henkilötietoja)
- Säilytysajan pituus, säilytysajan peruste
- Julkisuusluokka (liittyy käyttörajoituksiin)
 - Salassapitoaika, salassapidon peruste, suojaustaso, turvallisuusluokka
 - Käyttäjryhmä (metatiedot pakollisia vain salassa pidettäville)
- Identifiointitunnus
- Luontipäivämäärä (avattu/laadittu)
- Kieli (kielikoodi)
- Nimeke (kuvailtavan kohteen varsinainen nimi)
- Toimija (henkilön nimi asiakirjalle pakollinen)
- Versio (jos asiakirjasta on useita versioita)

SÄHKE2:ssa määriteltyjä valinnaisia hakuun vaikuttavia metatietoja (Arkistolaitos, 2009, 6-29.):

- Aihe
- Kuvaus (eli kohteen sisällön kuvailu)
- Tiivistelmä
- Toimijan rooli ja yhteisö
- Kohteiden väliset viittaussuhteet ja korvaavuus
- Julkistamisajankohta
- Muokkauspäivämäärä
- Sisällön kuvaus
- Kattavuus

- Hallinnon ala
- Alueellinen kattavuus
- Ajallinen kattavuus

SÄHKE2:n metatietomallissa esimerkiksi kuvailtavan kohteen aihe on valinnainen, eli sitä ei ole pakko tallentaa (Arkistolaitos, 2009a, 6). SÄHKE2-määräyksen metatiedoissa on runsaasti elinkaarihallintaan liittyviä metatietoja, koska se on tarkoitettu ohjeeksi organisaatioiden asiakirjallisen tiedon hallintaan (Arkistolaitos, 2009b, 3).

2.3. Bibliografinen kuvailu, FRBR käsitteellisenä mallina ja RDA-säännöstö

Kirjastoaineistoja kuvaillaan ja luetteloidaan niiden haettavuuden ja löydettävyyden parantamiseksi. Luettelointia tehdään lisäksi kulttuurisista syistä, esimerkiksi teosten rekisteröimiseksi (Suominen, Saarti, & Tuomi, 2009, 16). Kuvailu pyritään pitämään myös mahdollisimman yhdenmukaisena ja systemaattisena löydettävyyden takaamiseksi. Tässä käytetään apuna erilaisia kuvailustandardeja ja -ohjeita. (Suominen et al., 2009, 19.)

Luettelotietokantojen hakuelementit standardisoidaan, jotta haku voidaan kohdistaa tiettyyn teosta kuvaavaan tietoelementtiin ja niiden avulla voidaan tuottaa loogisesti järjestyviä lineaarisia luetteloita kuten selailunäkymiä (Suominen et al., 2009, 53). Myös *auktoriteettitiedostojen* ja -tietueiden tarkoituksena on hakuominaisuuksien parantaminen. Auktoriteettitiedoston tietueisiin tallennetaan tietyn käsite- tai ilmaisuryhmän käsitteiden erilaisia muotoja ja niihin merkitään mikä muodoista on auktorisoitu ja mitkä variantteja. Auktoriteettitiedosto yhdistää myös sisällönkuvailuina käytettävät ilmaisut ja niiden muodot ja lisäksi niiden avulla voidaan erottaa esimerkiksi keskenään samannimisten henkilöiden tiedot. (Suominen et al., 2009, 53-54)

Auktoriteettitiedostojen merkitys korostuu *FRBR-mallissa* (Functional Requirements for Bibliographic Records). Malli kuvaa luetteloinnin yleisimpiä käsitteellisiä lähtökohtia ja sen tavoitteena on soveltaa kuvailuperiaatteita tietotekniikan ja internetin tarjoamiin mahdollisuuksiin. (Suominen et al., 2009, 20, 55.) FRBR on entiteetti-relaatio-malli, jossa kuvataan bibliografista universumia, johon sisältyy entiteettejä (olioita) ja niiden välisiä suhteita (Suominen et al., 2009, 55). Entiteetit on jaettu kolmeen ryhmään: teos, jonka olomuotoja ovat myös ekspressio, manifestaatio ja kappale; henkilöt ja yhteisöt, jotka ovat vastuussa teoksesta sekä käsite,

tapahtuma, objekti ja paikka, jotka viittaavat dokumentin sisältöön. Kuvattavalla kohteella on siis erilaisia olomuotoja. Teos voi ilmetä ekspressioiden, manifestaatioiden ja kappaleina ja se on itsessään melko abstrakti käsite. Ekspressio on esimerkiksi teoksen toteutus toisessa ilmaisumuodossa. Manifestaatio on joukko, jonka kaikki kappaleet ovat keskenään samanlaisia ja kappale on yksi konkreettinen teoskappale. (Suominen et al., 2009, 56-57.)

FRBR:n entiteetit ovat (Coyle, 2010, 21):

- ryhmä 1: aineistonkuvailuentiteetit
 - teos
 - ekspressio
 - manifestaatio
 - kappale
- ryhmä 2: tekijöidenkuvailuentiteetit
 - henkilö
 - yhteisö
 - suku
- ryhmä 3: aiheenkuvailuentiteetit
 - käsite
 - paikka
 - tapahtuma
 - esine.

FRBR tulisi Coylen (2010, 23) mukaan nähdä entiteettien suhdeverkoston mallina, ei pelkästään yksittäisen bibliografisen teoksen kuvailuna. Entiteetti-relaatiomallin avulla voidaan esimerkiksi muodostaa useammanlaisia toisistaan poikkeavia näkymiä bibliografisiin tietoihin. (Coyle, 2010, 23.) Myös van Ballegooie ja Borie (2014, 76-77) näkevät aineistojen väliset suhteet tärkeinä. He ovat tarkastelleet FRBR:a ja linkitettyä dataa sarjajulkaisukatalogien sisällön löydettävyyden näkökulmasta. He ovat todenneet, että hakujärjestelmät ovat kehittyneet verkkoteknologian myötä, mutta uudemmatkaan järjestelmät eivät palvele kaikkia hakutarpeita. Tämä johtuu siitä, että järjestelmät eivät kuvaa riittävästi aineistojen välisiä suhteita. (van Ballegooie & Borie, 2014, 77.) MARC-formaatin tarkoitus oli alun perin olla vain luetteloinnin apuväline kirjastojen välillä, eikä se sovellu kovin hyvin hakujärjestelmiin eikä FRBR:n vaatimuksiin aineistojen välisten suhteiden kuvaajana (van Ballegooie & Borie, 2014, 78).

Nimer (Nimer, 2010, 227-228) kirjoittaa, että yksi tapa yhdistää kirjasto- ja arkistoaineiston kuvailu on ollut *MARC*-formaatin (*MAchine-Readable Cataloging*) käyttäminen myös arkistoaineiston kuvailussa. Ensisijaisten ja toissijaisten aineistojen sijoittaminen samaan tietokantaan onkin edistänyt aineistojen saatavuutta. Tämä ei kuitenkaan ratkaissut kaikkia aineistojen erilaisuudesta johtuvia saatavuusongelmia. Tällä hetkellä metadatarakenteiden ja kuvailustandardien yhdistämistä tehdään *RDA*:n (*Resource Description and Access*) kehityksen kautta. (Nimer, 2010, 227-228.) *RDA* on uusien teoreettisten lähtökohtien mukainen kirjastojen luettelointisäännöstö, jonka on tarkoitus korvata edellinen *AACR2-säännöstö* (*Anglo-American Cataloguing Rules*). *RDA*:n avulla pyritään laajentamaan kuvailtavien materiaalien joukkoa ja siihen on otettu mukaan myös arkistokuvailua hyödyttäviä muutoksia. (Nimer, 2010, 228.)

Keenanin (2014, 464) mukaan *RDA* on joustavampi standardi verrattuna *AACR2*:een, koska sitä voivat käyttää myös muut informaatioyhteisöt kuin kirjastot. Sen kehittäjät odottavat sen myös poistavan kulttuuriset erot kuvailussa ja tarjoavan ohjeistusta digitaalisten aineistojen kuvailuun sekä tehostavan käyttäjien haku- ja selailumahdollisuuksia. (Keenan, 2014, 464.) Greg Bakin (2012) mukaan nykyiset kuvailusäännöt ja asiakirjojen luokittelusäännöt on rakennettu palvelemaan enemmän arkistojen tarpeita kuin asiakirjojen luojien tai niiden käyttäjien tarpeita. Hän tarkastelee luokittelua elektronisten aineistojen hallinta järjestelmien (*Electronic Records Management System, ERMS*) kautta. Bak kritisoi tehtäväpohjaisen luokittelun näkemistä ainoaksi oikeaksi luokittelutavaksi nykyisessä arkistotutkimuksessa ja -diskurssissa. (Bak, 2012, 290-291.)

RDA:n käsitteellisinä lähtökohtina ovat *FRBR* ja *FRAD* (*Functional Requirements for Authority Data*) (Nimer, 2010, 230). *FRBR*:n käsitteellinen malli ei kuitenkaan sovi arkistomateriaaliin kaikkien mielestä esimerkiksi aineiston ainutkertaisen luonteen vuoksi (Nimer, 2010, 234). Myös teos-käsitteen sopivuudesta arkistomateriaaliin on epäselvyyttä. Thurmanin mukaan *FRBR*:llä on vain vähän vaikutusta arkistokuvailuun, koska *FRBR*:n avaintiteetti eli teos ei voi edustaa arkiston organisoinnin keskeistä yksikköä (Thurman, 2007, 97). Arkistomateriaalin keskeinen yksikkö on Thurmanin mukaan kokoelma tai fondi, mutta kumpikaan näistä termeistä ei ole *FRBR*:n entiteetti (Thurman, 2007, 98). Arkistoaineiston jokaisella teoksella voisi Thurmanin mukaan olla vain yksi ekspressio tai manifestaatio, minkä vuoksi käsitteellisesti monitasoinen kuvailu ei olisi arkistojen kohdalla mielekästä (Thurman, 2007, 99). Nimer kuitenkin huomauttaa, että myös suurimmalla osalla muitakin aineistotyypppejä on vain yksi manifestaatio. Teos-käsitteen hyöty tulee Nimerin mukaan esille entiteettien välisten suhteiden kuvaamisessa. Sekä osa- että kokosuhteet voidaan Nimerin mukaan kuvata teostasolla. (Nimer, 2010, 234.)

RDA:n suunnittelussa on ajateltu aineistojen käyttäjien näkökulmaa enemmän kuin aikaisemmissa standardeissa. RDA:n tarkoituksena on kuvailla aineistoja siten, että se helpottaa aineiston käyttäjän tehtäviä, joihin kuuluu löytäminen, tunnistaminen, valinta ja aineiston käyttöön saaminen. RDA perustuu FRBR-käsittemalliin ja se on suunniteltu toimimaan jo käytössä olevien kuvailuformaattien (kuten MARC ja XML) ja myös tulevien rakenteellisten kuvailuformaattien kanssa. Se myös helpottaa aineistojen ja henkilöiden välisten suhteiden kuvailua. (Keenan, 2014, 447.) Suhteet kuvataan RDA:lla tietokoneiden ymmärtämässä rakenteellisessa muodossa. Myös henkilöiden tai organisaatioiden rooli tallennetaan rakenteellisessa muodossa, mikä osaltaan helpottaa käyttäjän tehtäviä. (Keenan, 2014, 459.) Yksi tapa kuvata aineistojen ja ihmisten välisiä suhteita on tietojen tallentaminen auktoriteettitietueisiin. RDA tarjoaa myös tähän ohjeistusta. (Keenan, 2014, 459.)

RDA:n myötä tehtyjen muutosten on tarkoitus parantaa kuvailevien luetteloiden saavutettavuutta. Yksi muutoksista on lyhenteiden käytön vähentäminen. RDA sisältää myös kuvailussa käytettyjen termien laajemman standardisoinnin (Nimer, 2010, 234). Arkistoaineiston kuvailuun vaikuttavat muutokset perustuvat pääasiassa ydinstandardin (core standard) vaatimiin lisäelementteihin (Nimer, 2010, 236). Ydinlementtivaatimukset RDA-perusteisessa luetteloinnissa vaihtelevat aineistotyyppin mukaan. RDA tarjoaa myös erilaisia vaihtoehtoja kuvailuun, niitä voi soveltaa luetteloijan harkinnan tai instituution käytännön mukaisesti. (Nimer, 2010, 235.) RDA sisältää kuvailuvaatimuksia myös henkilöiden, perheiden ja organisaatioiden kuvailuun. Nämä tallennetaan useimmiten attribuutteina riippumatta siitä, käytetäänkö niitä hakemisen apuna vai ei (Nimer, 2010, 236).

2.4. Kulttuuriperintöaineistojen kuvailun yhdistäminen ja AHAA-käsittemalli

Suomessa on pohdittu arkistojen, kirjastojen ja museoiden käytössä olevien tietojärjestelmien yhteiskäytön edistämistä jo 1990-luvulta saakka. Yhteiskäytön vaatimuksena on jonkinasteinen yhteinen tietorakenne, minkä vuoksi aineistojen kuvailussa ja luetteloinnissa tarvitaan muutoksia yhteistyössä eri muistiorganisaatioiden välillä. (Vakkari, Simovaara, & Valanto, 2004, 3.) Yhteistyöhön liittyvä Kamut-hankkeen työryhmä päätyi loppuraportissaan suosittelemaan yhteisen metadataformaatin käyttöönottoa. Jo Kamut-hankkeessa yhteiset auktoriteettitiedot ainakin henkilö- ja yhteisönimien kohdalla nähtiin haun kannalta oleellisena aineistojen hakua parantavana tekijänä. (Kamut, 1997, 5.)

Kansallinen digitaalinen kirjasto (KDK) on arkistojen, kirjastojen ja museoiden laaja yhteinen kokonaisarkkitehtuurihanke. Arkistojen AHAA-hanke kytkeytyy KDK-hankkeeseen ja AHAA:ssa tehtävä käsittemallikehitys tulee vaikuttamaan mm. arkistohakemistopalveluun, jonka asiakaskäyttöliittymänä toimii Finna. Finna on tarkoitettu arkisto-, kirjasto- ja museoaineiston tiedonhaun portaaliksi (Finna, 2013). Finnan tarkoituksena on tulla tutkimuksen, opetuksen ja oppimisen ydinpalveluksi, jonka kautta löytyvät monenlaiset tietosisällöt erilaisiin ja muuttuviin tarpeisiin (Hormia-Poutanen, 2013). Finnaan on suunnitteilla arkistosektorin oma näkymä eli käyttöliittymä. Finna on web 2.0 -sovellus, joten se tarjoaa esimerkiksi kommentointimahdollisuuden ja muita vuorovaikutteisia toimintoja. AHAA-hankkeessa toteutettavaan arkistosektorin omaan Finna-näkymään on suunnitteilla mahdollisuus osallistua aineistojen kuvailuun. Tämän tarkoitus on parantaa aineiston löydettävyyttä. AHAA-hankkeessa on tällä hetkellä mukana Arkistolaitos, Suomalaisen kirjallisuuden seura, Porvarillisen työn arkisto, Keskustan ja maaseudun arkisto, Svenska centralarkivet ja CSC - Tieteen tietotekniikan keskus (Arkistolaitos, 2014).

Suomessa kehitetään arkistokuvailulle käsitteellistä mallia KDK:n (Kansallinen digitaalinen kirjasto) määrittämässä viitekehyksessä. KDK suosittaa, että kaikki kulttuuriperintöaineiston kuvailu pohjautuu FRBR:iin perustuvaan kirjastokuvailustandardiin RDA:iin (Resource Description and Access). Uusi arkistokuvailun kansallinen käsitemalli ottaa huomioon myös ICA-standardit sekä uudelleen määritellyn proveniensiiperiaatteen. Käsitemallissa on oleellista kuvailuentiteettien erillisyys ja niiden keskinäiset suhteet. Digitaalisessa toimintaympäristössä aineiston intellektuaalisen hallinnan merkitys kasvaa, eikä tietovarantojen hallintaa voi enää perustaa aineiston fyysiselle järjestämiselle. (Henttonen & Kilkki, 2013, 5, 10-11.) Proveniensi on uuden tulkinnan mukaan ”*käsitteellinen entiteetti ja multilateraalinen kontekstien suhdeverkosto*” (Henttonen & Kilkki, 2013, 11). Provenienssin kuvaamiseen monitahoisina suhteina vaaditaan uudenlainen kuvailutapa (Henttonen & Kilkki, 2013, 14). Aineiston metatiedot koostuvat arkistokuvailussa aineistoa kuvailevasta metatiedosta ja aineiston provenienssia kuvailevasta metatiedosta (Henttonen & Kilkki, 2013, 23). AHAA-hankkeessa ne on nimetty tietovarantoentiteeteiksi ja kontekstientiteeteiksi (Henttonen & Kilkki, 2013, 24). AHAA-käsitemalli ei määrittele aineiston kuvailua käytännössä. Käsitemallia käytetään varsinaisten kuvailusääntöjen ja metatietomallien pohjana. (Henttonen & Kilkki, 2013, 23.)

Arkistokuvailun kansallisen käsittemallin luonnoksessa (Henttonen & Kilkki, 2015) kuvailuentiteetit on nimetty luokittain seuraavasti:

- Toiminta
 - Tehtävä
 - Elinkaarihallinta
- Valtuutus
- Toimija
- Paikka
- Tapahtuma (representoi toimijan toimintakontekstiin liittyvää ajallista ilmiötä)
- Aineisto
 - Manifestaatio
- Aihe

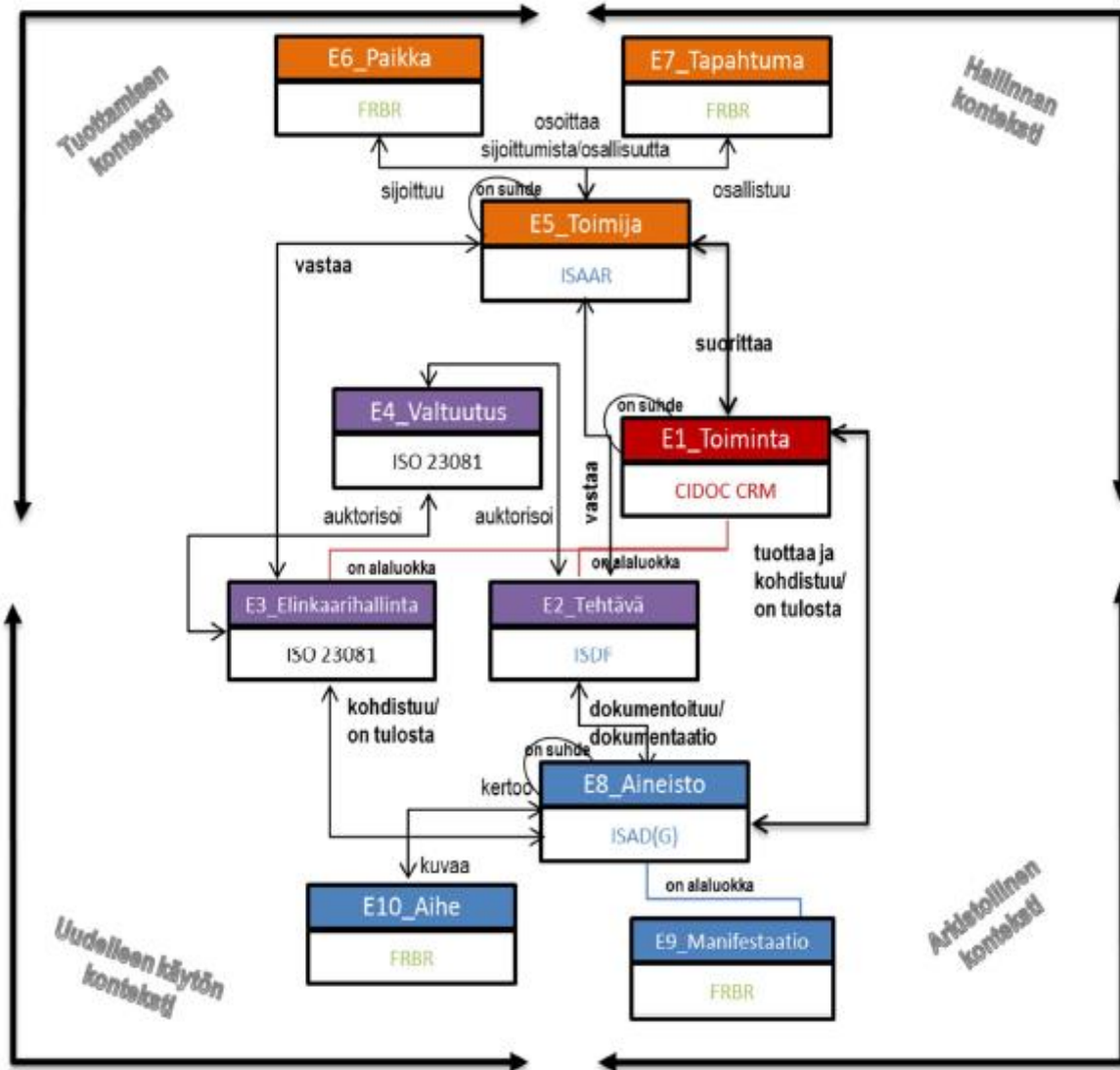
Taulukko 4: Käsittemallin entiteetit ja niiden väliset suhteet (Henttonen & Kilkki, 2015, 29).

	E1_ Toiminta	E2_ Tehtävä	E3_ Elinkaari- hallinta	E4_ Valtuutus	E5_ Toimija	E6_ Paikka	E7_ Tapahtuma	E8_ Aineisto/ E9_ Manifestaatio	E10_ Aihe
E1_ Toiminta		Ylänokka	Ylänokka		1a_ suorittaa			1b_tuottaa ja kohdistuu	
E2_ Tehtävä	Aläknokka			2a_ ankkuroisoi	2b_vastaa			2c_dokumentoituu	
E3_ Elinkaari- hallinta	Aläknokka			3a_ ankkuroisoi	3b_vastaa			3c_kohdistuu	
E4_ Valtuutus		4a_ ankkuroisoi	4b_ ankkuroisoi						
E5_ Toimija	5a_ suorittaa	5b_ vastaa	5c_ vastaa			5d_ sijoittuu	5e_ osallistuu		
E6_ Paikka					6a_osoittaa sijoittumista				
E7_ Tapahtuma					7a_osoittaa osallisuutta				
E8_ Aineisto/ E9_ Manifestaatio	8a_tulosta	8b_dokumentaatio	8c_tulosta						8d_kertoo
E10_ Aihe								10_a kuvaava sisältöä	

Sama käsitteellinen sisältö voi esiintyä kahtena tai useampana eri materiaalisena objektina sekä arkistossa että kirjastossa esimerkiksi mikrokuvauksen seurauksena. Arkistokuvailussa tarvitaan FRBR-käsittemallin ensimmäisestä ryhmästä kuitenkin vain käsitteitä ekspressio (uudemmassa versiossa aineisto) ja manifestaatio Henttonen ja Kilkin (2013, 22) mukaan. Ne representoivat tietovarantoa arkistokuvailussa, ekspressio (aineisto) tekee sen käsitteellisellä tasolla ja manifestaatio edustaa tietovarantoa materiaalisena ja konkreettisena objektina (Henttonen & Kilki, 2013, 22).

FRBR:lle ja arkistokuvailun kansalliselle käsitemallille yhteisiä entiteettejä ovat toimija, paikka, ajallinen ilmiö ja aihe. Myös aineistoekspressio (aineisto) ja -manifestaatio-käsitteet juontuvat FRBR-käsitemallista (Henttonen & Kilki, 2013, 22). Oleellisinta uudistuksessa ja molemmissa käsitemalleissa ovat kuitenkin entiteettien väliset suhteet, jotka järjestelmissä toteutetaan linkittämällä auktoritietueissa sijaitsevat entiteetit toisiinsa (kuva 1). Suhteiden kuvailu vaatii tietovarantoentiteettien ja kontekstientiteettien erottamista toisistaan (taulukko 4). Myös kuvailusääntöjen ja metatietomallien täytyy olla modulaarisia, jotta suhteet voidaan luoda ja kuvailla. (Henttonen & Kilki, 2013, 24.)

Arkistoaineistoon liitettävissä metatiedoissa tarvitaan kuvailutietojen lisäksi myös elinkaarihallintaan liittyviä metatietoja (kuva 1) (Henttonen & Kilki, 2013, 9). Elinkaarihallintaan liittyvä entiteetti on ilmeisten (elinkaaritapahtuma-entiteetit) lisäksi esimerkiksi myös toimija, sillä tietovarantoon kohdistuvat toimenpiteet ja niiden suorittajat on hyvä tuoda esille muun muassa toiminnan läpinäkyvyyden varmistamiseksi (Henttonen & Kilki, 2013, 10).



Kuva 1: Arkistokuvailun käsittemalli ja elinkaaren jatkumon kontekstit (Henttonen & Kilkki, 2015, 13).

Käsitteellisen tason kuvailu-uudistusta tehdään myös kansainvälisesti. International Council on Archives (ICA) on nimennyt asiantuntijaryhmän kehittämään kattavan kuvailustandardin, joka perustuu neljälle jo olemassa olevalle erilliselle standardille (mainittu luvussa 2.1 sivulla 7). Kehitystyö on määritelty tehtäväksi käsitteellisen mallintamisen menetelmillä. (Gueguen, Manoel, Pitti, & Grimoüard, 2013, 568.) Asiantuntijaryhmän eli Experts Group on Archival Description (EGAD) työ alkoi vuonna 2012 ja se on suunniteltu valmistuvaksi 2016 (Gueguen et al., 2013, 578). EGAD:n tavoitteena on saada kansainvälinen hyväksyntä uudelle kuvailumallille ja mallin tulee olla sovitettavissa kaikkiin kulttuureihin ja kieliin. (Gueguen et al., 2013, 579.) Aineistokuvailun uudistaminen on jo pitkään painottunut kuvailun avainkomponenttien erottamiseen ja yhdistämiseen niiden välisten suhteiden kautta saatavuuden parantamiseksi. Tämä vähentää kuvailutyön päällekkäisyyttä sekä parantaa kuvailun täsmällisyyttä ja kuvailtujen aineistojen ymmärrettävyyttä. (Gueguen et al., 2013, 581.)

3. KATSAUS TIEDONHAKUTUTKIMUKSEEN JA -KIRJALLISUUTEEN

Tässä luvussa esitellään tutkimuksia, joissa on pyritty selvittämään tiedonhakijoiden näkökulmia ja heidän tietotarpeitaan sekä hakustrategioitaan. Alaluvussa 3.1 kerrotaan tiedonhausta yleisesti ja sen muuttumisesta internetin käytön yleistymisen myötä. Arkistonkäyttäjien tiedonhakutarpeistakin on tehty jonkin verran tutkimusta, mutta ne ovat keskittyneet usein vain johonkin yksittäiseen arkistotiedonhaun ongelmaan. Näistä tutkimuksista kerrotaan alaluvussa 3.2. Luvussa 3.3 kerrotaan linkitetystä datasta ja semanttisen tiedonhaun tukena toimivista ontologioista. Tässä luvussa kerrottujen tutkimuksien tuloksia käytetään hyväksi hakukriteerilistan muodostamisessa.

3.1. Tiedonhankinta ja tiedonhaku

Informaation hankinta kuvataan yleensä prosessina, joka voidaan jakaa eri vaiheisiin. Prosessiin sisältyy yleensä ainakin ongelman ja tiedontarpeiden määrittäminen, tiedonlähteen valinta, relevantin osuuden valinta ja tiedon käyttö. Tiedonhankinta on useimmiten iteratiivista, eli eri vaiheet tapahtuvat useamman kerran prosessin aikana. (Haasio & Savolainen, 2004, 43-45.) Tiedonhaku sähköisten järjestelmien kautta on osa tiedonhankintaa (Haasio & Savolainen, 2004, 37). Tiedonhakutoimintakin voidaan jakaa eri vaiheisiin. Alaterä ja Halttunen (2002, 84) ovat kuvanneet tiedonhaun käsittävän 11 vaihetta. Niihin kuuluu muun muassa hakukäsitteiden ja hakutermien ideointi ja valinta sekä hakutermien yhdistely käytettävän tiedonhakujärjestelmän hakukielen mukaisesti. Tiedonhakutapa on kuitenkin aina riippuvainen myös tiedonhaun ja tiedon käytön tarkoituksista. (Alaterä & Halttunen, 2002, 84-85.)

Kuvailutietokantajärjestelmiin tallennetaan erilaisia tietoja. Osa järjestelmistä mahdollistaa haun luonnollista kieltä käyttämällä. Luonnollisen kielen mukainen kuvailu voidaan tallentaa esimerkiksi avainsanoina, jotka kuvaavat aineiston tai dokumentin sisältöä tai tiivistelmänä dokumentista tai aineistosta. (Alaterä & Halttunen, 2002, 69-70.) Luonnollisen kielen käyttö vaatii hakujärjestelmältä edistyneitä hakua käsitteleviä toimintoja. Dokumentaatiokielet eli kontrolloidut sanastot määrittelevät kuvailijoille ja tiedonhakijalle yhteisen kielen. Tämän on tarkoitus auttaa hakijaa löytämään etsimänsä kuvailutietojärjestelmistä. (Alaterä & Halttunen, 2002, 71.) Tietokantoihin voidaan tallentaa myös dokumentin teksti kokonaisuudessaan, jolloin haku voidaan kohdistaa siihen (Alaterä & Halttunen, 2002, 68).

Tiedonhaussa olennaista ovat hakutulokset ja niiden relevanssi. Hakutulos on relevantti silloin, kun hakutulokset vastaavat hakijan tiedontarpeita. Alaterä ja Halttunen (2002, 126) ovat

määritelleet kaksi erilaista relevanttiuden muotoa, aiherelevanssin ja käyttäjärelevanssin. Aiherelevanttius toteutuu silloin, kun hakukysymys ja dokumentin kuvaus täsmäävät. Käyttäjärelevanssi ottaa huomioon myös käyttäjän tehtävän ja muun kontekstin. Hakutuloksissa saattaa olla esimerkiksi käyttäjälle jo tuttuja dokumentteja, jolloin ne eivät vastaa hakijan tiedontarpeita. (Alaterä & Halttunen, 2002, 126.)

Jennifer Schaffner (2009) kirjoittaa, että arkistojen näkymättömyys voi johtua enemmän tuotetusta metadatasta kuin aineistojen hakuun tehdyistä käyttöliittymistä. Tämä voi johtua siitä, että ihmiset olettavat löytävänsä arkistot ja niiden kokoelmat kattavasti avoimesta verkosta samoilla tekniikoilla kuin muutkin haettavat tiedot. (Schaffner, 2009, 4.) Uusia mahdollisuuksia päästä informaation äärelle tarvitaan nykyisin yleisen oletuksen vuoksi, jonka mukaan kaikilla pitää olla pääsy kaikkeen informaatioon. Asiakirjatkaan eivät ole enää valtion omaisuutta siinä määrin kuin ennen. (Huvila, 2008, 17.) Käyttäjät olettavat kaikkien elektronisten aineistojen olevan heidän käytettävissään heidän haluamallaan tavalla (Adams, 2007, 31).

Tiedonhaku on muuttunut internetin ja hakukoneiden yleistymisen myötä. Georgaksen (2013; 2014) tutkimusten mukaan nykyiset opiskelijat pitävät itseään hyvinä tiedonhakijoina ja he käyttävät usein Googlea tiedonhaussa. Google ja Google Scholar ovat tehneet helpoksi löytää informaatiota mistä tahansa tutkimusaiheesta. Kirjastojen hakujärjestelmissä tarvitaan erilaisia hakutapoja kuin Googlea käytettäessä. (Georgas, 2014, 503-504.) Hän luokitteli hakukyselyt kolmeen kategoriaan, jotka olivat luonnollisen kielen kyselyt, yksinkertainen avainsana- tai fraasikysely ja Boolean-kysely. Googlella tehdyistä hauista yleisimmin tehtiin hakuja luonnollisen kielen kyselyn ja avainsanan tai fraasin yhdistelmällä. Vain yksi tutkittavista (yhteensä 29 kpl) käytti kaikkia kolmea hakukyselytapaa. Boolean-operaattoreita käytti viisi tutkittavasta henkilöistä. Yhdistelmähakuvälineellä (federated search tool) tehtiin yleensä yhdistelmiä erilaisista kyselytavoista. Suosituin oli yhdistelmä avainsana- tai fraasikysely ja luonnollisenkielenkysely. Boolean-operaattoreita ei käytetty kovin edistyneesti yhdistelmähakuvälineelläkään. (Georgas, 2014, 509.) Yhdistelmähakujärjestelmässä käytettiin kuitenkin enemmän Boolean-hakukyselyä kuin Googlea käytettäessä, mikä puhuu sen puolesta, että mahdollisuus siihen kannattaa tarjota jo ensimmäisellä hakusivulla (Georgas, 2014, 511). Tässä käsitelty artikkeli ei kuitenkaan vielä kerro, kuinka tehokkaita eri hakukyselymenetelmät olivat, vaan tulokset siitä julkaistaan myöhemmin (Georgas, 2014, 509).

Tutkimustulosten mukaan opiskelijoilla oli melko huonot tiedonhakutaidot (Georgas, 2014, 527). Tutkittavat opiskelijat eivät esimerkiksi käyttäneet juurikaan aikarajausta hauissaan, vaikka osassa

tiedonhakuaiheista se olisi ollut hyvin oleellista ja viimeisimmän tutkimustiedon löytäminen on oleellista kaikessa akateemisessa tutkimuksessa (Georgas, 2014, 512). Google on hakukoneena hyvä, koska se on joustava ja anteeksiantava sekä reagoi monenlaisiin toimintoihin. Kirjastojen tarjoamia hakuvälineitäkin kuitenkin tarvitaan, koska ne ovat luotettavampia. Hakuvälineiden tulisi kuitenkin kyetä auttamaan käyttäjää jokaisessa informaationhakuprosessin vaiheessa. Tutkija tuli siihen tulokseen, että hyvän tuloksen aikaansaamiseksi tarvitaan sekä Google-tyyppistä hakua että tarkemman haun mahdollistavia välineitä. (Georgas, 2014, 527-528.)

Myös aktiivisesti arkistojen aineistoja käytävillä historioitsijoilla on puutteita digitaalisessa tiedonhaussa. He kertoivat haastattelussa, etteivät mielellään käytä esimerkiksi tarkennettua hakua ja raportoivat itse tietojenkäsittelytaitonsa alle keskitason. (Elena, Katifori, Vassilakis, Lepouras, & Halatsis, 2010, 33.)

3.2. Arkistonkäyttäjien informaatiotarpeet ja hakukäyttäytyminen

Tutkimusten mukaan käyttäjät eivät yleensä halua etsiä kokoelmia provenienssin perusteella, joten rakenteellinen metadata ei käyttäjien kannalta ole välttämättä paras mahdollinen vaihtoehto, vaikka rakenteellinen metadata on arkistokokoelmien hallinnassa erittäin tärkeä väline (Schaffner, 2009, 4). Huvilan haastattelemat asiantuntijat olivat sitä mieltä, että käyttäjillä on useita erilaisia tiedonhakustrategioita käytössään. Asiantuntijoiden mukaan löydettävyyys ja saavutettavuus ovat tärkeimmät arkiston toiminnot ja internet tarjoaa tähän hyvän väylän. (Huvila, 2008, 21.) Jennifer Bunnin (2013) mukaan kuvailun ensisijainen tarkoitus on kuitenkin autenttisuuden ja merkityksen osoittaminen, mikä on vaarassa jäädä informaation jaon jalkoihin (Bunn, 2013, 244). Hän korostaa, että useat uudistusaskleet arkistokuvailustandardien ja -käytäntöjen saralla eivät ole onnistuneet kohdentamaan huomiota sekä arkistokuvailun sisäiseen merkitykseen arkistoaineistomaailmassa että informaation jakamiseen muille eli ulkoiseen merkitykseen, vaan ainoastaan toiseen näistä (Bunn, 2013, 241).

Kokoelmien sisältö on käyttäjien kannalta olennaisempaa kuin sen muoto. Aineiston haku aiheen perusteella on käyttäjille eri tutkimusten mukaan tärkeää. (Schaffner, 2009, 6.) Käyttäjät saattavat tyytyä tunnettujen aineistojen hakuun, mikäli aiheen mukaista pääsyä kokoelmaan ei tarjota (J. Dooley, 1990 Schaffnerin, 2009, 6 mukaan). Lisäksi avainsanoilla tehtävä haku, esimerkiksi ihmisten nimillä ja paikannimillä tehtävä haku on käyttäjille, etenkin humanististen alojen tutkijoille, tutkimusten mukaan tärkeää (Schaffner, 2009, 7). Käyttäjät haluavat siis käyttää monenlaisia hakumenetelmiä, myös esimerkiksi selata aiheita tai avainsanoja. Tämä on

ongelmallista sen vuoksi, että arkistoteoria ja käytännöt ovat ohjanneet kuvailua eri suuntaan, kuvailuissa kerrotaan yleensä, mistä kokoelma koostuu. (Schaffner, 2009, 8.)

Monissa arkistojen käyttäjiin kohdistuneissa tutkimuksissa arkistonkäyttäjät on eroteltu erilaisiin ryhmiin. Näitä ryhmiä ovat muun muassa historioitsijat, sukututkijat ja akateemiset tutkijat esimerkiksi humanistisilta aloilta (Duff & Johnson, 2003; Lindquist, Dulock, Törnroos, Hyvönen, & Mäkelä, 2013; Adams 2007.) Hakukäyttäytymiseen vaikuttavat roolin ja tehtävän lisäksi myös saatavilla olevat hakujärjestelmät ja niiden toimintaperiaatteet (Duff & Johnson, 2003; Hamburger, 2004).

Adams (2007, 27) on nimennyt kaksi arkistonkäyttäjärühmää, joista ensimmäinen ryhmä on alkuperäistutkimuksen tekijöitä. Alkuperäistutkimuksen tekijät kattavat suurimman osan akateemisista tutkijoista. Toinen hänen ryhmistään koostuu sellaisista tiedonhakijoista, jotka käyttävät arkistoaineistoa lähteenä mille tahansa faktuaaliselle tai henkilökohtaiselle informaatiolle. Jälkimmäisen ryhmän tutkimus on pienimuotoisempaa ja käytetyt lähteet ovat keskenään samankaltaisia. Ensimmäisen ryhmän jäsenet etsivät useammanlaisia asiakirjoja seuraten yksilöllisesti määritetyltä tavoitteita. (Adams, 2007, 27.)

Colorado Boulderin yliopistossa tutkittiin humanististen alojen opiskelijoiden ja henkilökunnan hakutarpeita alkuperäisaineistoja haettaessa. Tutkijoiden mukaan tuloksia voidaan jossain määrin soveltaa kaikkiin alkuperäislähteitä verkossa etsiviin käyttäjiin. Lindquistin ja Longin (2011) suorittamassa tutkimuksessa haluttiin selvittää humanististen alojen opiskelijoiden nykyistä teknologian käyttöä sekä sitä, miksi verkossa olevia alkuperäisaineistoja käytetään opiskelussa tällä hetkellä melko vähäisesti. (Lindquist & Long, 2011, 225.)

Tulosten mukaan sekä opiskelijoilla että henkilökunnan jäsenillä oli edelleen vaikeuksia löytää ja kontekstualisoida verkosta arkistoaineistoja, eivätkä he olleet tietoisia kaikista saatavilla olevista aineistoista. Myös kokoelman sisällä olevien yksittäisten lähteiden ja niiden sisältämän informaation löytäminen tuotti ongelmia. Haastatellut raportoivat bibliografisen metadatan olevan usein riittämätöntä yksittäisten lähteiden löytämiseen, mikä johtuu lähteisiin sisältyviin aiheen, ajan ja maantieteellisen alueen liian yleisellä tasolla olevasta kuvailusta. (Lindquist, Dulock, Törnroos, Hyvönen, & Mäkelä, 2013 915.)

Avainsanahaku on Lindquistin ja muiden kirjoittajien mukaan (2013, 915) sattumanvaraista, sillä samantapaisia käsitteitä on käytetty monin eri tavoin eri teksteissä ja jopa tekstien sisällä. Avainsanahaku rohkaisee käyttäjiä myös katsomaan kirjoittajien mukaan vain dokumenttien

katkelmia (snippet), mikä vähentää dokumenttien kontekstin ymmärtämistä. (Lindquist et al., 2013, 915-916.) Arkistoaineistojen kontekstualisointi on tarpeellista monien tutkijoiden mukaan, koska se auttaa käyttäjiä ymmärtämään paremmin aineiston olennaisimman sisällön. (Lindquist et al., 2013, 916). Esimerkiksi historiallisen kontekstin esiin tuominen auttaisi ymmärtämään asiakirjoja paremmin (Lindquist & Long, 2011, 233).

Duff ja Johnson (2003) haastattelivat sukututkijoita selvittääkseen heidän informaationhakukäyttäytymistään. Useimmat haastatelluista olivat sukututkimuksen asiantuntijoita, joten tulokset eivät välttämättä ole yleistettävissä harrastelijasukututkijoihin (Duff & Johnson, 2003, 82). Sukututkimus on luonteeltaan iteratiivista ja siinä käytetään useita erilaisia strategioita tiedon löytämiseksi (Duff & Johnson, 2003, 94). Haastateltujen mukaan sukututkimus ei ole vain nimien keräämistä ja yhdistämistä sukupuuhun, vaan siinä tarvitaan tietoa myös esimerkiksi paikoista ja tapahtumista henkilöiden löytymisen avuksi. Sukututkija tarvitsee siis taustatietoa tietyn ajankohdan tapahtumista ja esimerkiksi hallinnon toiminnasta löytääkseen lisäinformaatiota tietyistä suvusta. Tämän voi kuvata tapahtuvan kolmessa vaiheessa. (Duff & Johnson, 2003, 83-84.) Nimet ovat tärkeä hakuelementti sukututkijoille, mutta kaikki arkistojärjestelmät eivät tarjoa pääsyä aineistoihin nimihaun kautta. Tässä tapauksessa tarvitaan yleensä ensin tietoa sijainnista. Monet aineistot on järjestetty maantieteellisen sijainnin mukaan ja niihin on pääsy sen kautta, joten se on henkilöiden nimien jälkeen tärkein hakuelementti. Asuinpaikkakin voi kuitenkin olla hankala selvittää. (Duff & Johnson, 2003, 85-86.) Maantieteellisen sijainnin nimen määrittämiseksi sukututkijat tarvitsevat usein vanhoja karttoja, sillä rajat ja paikannimet muuttuvat pitkällä aikavälillä (Duff & Johnson, 2003, 94). Ajankohta on kolmas tärkeä hakuelementti sukututkijoille. Sukututkijat haluavat haussa yhdistää näitä hakuelementtejä tarkentaakseen hakuaan. (Duff & Johnson, 2003, 87.)

Kaikki tutkimukseen haastatellut tiesivät miten heidän informaatiotarpeensa tietyistä ihmisistä muunnettiin hakupyynnöksi koskien esimerkiksi tapahtumia. He olivat oppineet informaationhakutaitonsa sukututkimusta tehdessään. (Duff & Johnson, 2003, 88.) Aloittelevilla sukututkijoilla on vaikeuksia muuntaa tarpeensa saada ihmisistä tietoa oikeanlaisiksi hakupyynnöiksi. He tarvitsevat tietoa erityyppisten asiakirjojen sisällöstä ja luoista löytääkseen haluamaansa informaatiota. Useimmat järjestelmät vaativat käyttäjältä tietoa näistä asioista valmiiksi. (Duff & Johnson, 2003, 91.)

Elena, Katifori, Vassilakis, Lepouras ja Halatsis (2010) tutkivat historioitsijoiden arkistoissa tekemää tutkimusta. Heidän mukaansa hakuja on todella vaikea toteuttaa ilman kokeneen

arkistohenkilöstön tukea. Tutkijat tutkivat historioitsijoiden tiedontarpeita, hakustrategioita ja työskentelytapoja ottamalla selvää historioitsijoiden hakukyselyiden sisällöstä sekä haastatteleamalla heitä. Historioitsijoiden tavoitteena on luoda uudelleen menneisyys keräämällä ja yhdistelemällä historiallisia tietoja. (Elena et al., 2010, 25-26.) Tutkijat analysoivat ja luokittelivat noin sata luonnollisen kielen hakukyselyä. Hakukyselyistä suuri osa (42 %) koski joko henkilön tai organisaation biografiaa tai historiallista kehitystä. Kolmasosa hakukyselyistä kohdistui suoraan tiettyyn ajanjaksoon (epäsuorasti ajanjaksoon voi viitata esimerkiksi henkilön nimen avulla). Tutkijat päättelivät, että haetun asian kehityksen tuominen esille sekä mahdollisuus rajoittaa aikajaksoa on tärkeää historioitsijoiden työskentelyn tukemiseksi. Tutkijat tulivat myös siihen johtopäätökseen, että hakujärjestelmän tulisi tarjota tapoja paikallistaa dokumentteja niiden aiheen perusteella. Tämä voidaan toteuttaa monin tavoin, esimerkiksi avainsanoilla tai kontrolloiduilla taksonomioilla, joilla voi luokitella aihepiiriä tarkemminkin. (Elena et al., 2010, 28-29.)

Haastatteluissa historioitsijat kertoivat luottavansa enemmän perinteisiin tiedonhakumenetelmiin arkistoluetteloista kuin digitaaliseen hakuun. He käyttivät muun muassa harvempia avainsanoja tai niiden yhdistelmiä digitaalisessa haussa. Historioitsijat olettivat hakukoneen yhdistävän hakuun automaattisesti esimerkiksi synonyymit ja semanttisesti samaa tarkoittavat termit. Ongelmana digitaalisessa haussa koettiin myös epärelevanttien dokumenttien suuri määrä hakutuloksissa. Epärelevantit hakutulokset johtuivat suurimmaksi osaksi metadatan vähyydestä tai liian yleisellä tasolla olevasta metadatasta. Historioitsijat olivat huomanneet myös, että kaikki relevantit dokumentit eivät ole mukana hakutuloksissa. Tämä voi johtua monista eri syistä kuten luokittelutavasta, metadatan puutteesta ja sen epäjohdonmukaisuudesta. Lisäksi historiallisten arkistojen digitaalinen haku ei välttämättä tavoita dokumentin tarkoitusta, minkä vuoksi relevantteja dokumentteja saattaa puuttua haun tuloksista. (Elena et al., 2010, 32-33.)

Elena ja muut kirjoittajat (2010) esittävät suosituksia historiallisen tutkimuksen tueksi oman tutkimuksensa perusteella. Digitaaliseen aineistoon tulisi liittää historioitsijoiden tarvitsemia sisältöä kuvaavia metatietoja ja rakenteellista metatietoa. Historioitsijoille hyödyllistä metatietoa ovat etenkin dokumentin aihe, sisällön aiheen ajankohta, kirjoittaja ja dokumentissa viitattut entiteetit. Rakenteellinen metatieto tarkoittaa tässä sitä, että informaation tulisi olla varastoituna erillisinä ja liitettynä dokumenttiin. Pelkät avainsanat eivät riitä tässä yhteydessä tutkijoiden mukaan. Rakenteellinen metatieto lisää historioitsijoiden kontrollia hakuprosessissa. Kuvailun kustannusten vähentämiseksi kuvailua voidaan osittain automatisoida ja myös aineistojen käyttäjien voidaan antaa osallistua kuvailutiedon tuottamiseen esimerkiksi liittämällä kuvaavia

sanoja aineistoihin (tag) tai automaattisella käyttäjien toiminnasta saatavalla informaatiolla (Elena et al., 2010, 33-34.)

Tallennetut tiedot pitäisi tuoda esille taksonomiana tai semanttisena verkostona. Esimerkiksi aihe-metadata ja aikajana ovat tietoja, joiden esittäminen tietomallin avulla on hyödyksi aineistoja etsiville. Tietomalleihin tulisi sisällyttää sekä geneerisiä että spesifejä aineistoja kuvaavia käsitteitä. Myös historioitsijoiden oman käsitteellisen mallin ja aineiston kuvailuun käytetyn käsitteellisen mallin välisten eroavaisuuksien esiintuominen auttaa tutkijoita valitsemaan oikeat käsitteet hakuun. (Elena et al., 2010, 34.) Semanttisen tiedonhaun mahdollistava kuvailu ja kuvailujärjestelmä ovat hyödyksi tässäkin.

Hamburger (2004) yritti selvittää, miten tutkijat etsivät asiakirjakokoelmia arkistoista. Hän halusi tietää käyttävätkö he hakiessaan aineistoja avainsanoja, fraaseja, Boolean hakua, aihe-hakua, otsikko-hakua tai henkilön nimeä hakukriteerinä. Tutkimuksen kohteena oli myös se käyttävätkö tutkijat kontrolloituja sanastoja vai epätarkempia avainsanoja. (Hamburger, 2004, 82-83.) Tutkittavista 78 prosenttia suoritti avainsanahaun, 31 prosenttia teki haun käyttäen henkilön nimeä ja 23 prosenttia haki aiheen mukaisesti. Tutkittavat eivät kuitenkaan tutkijan pyynnön mukaisesti raportoineet hauissa käyttämiään termejä, joten niitä ei pystytty analysoimaan. (Hamburger, 2004, 84.) Hamburger suosittelee tutkimuksensa perusteella, että avainsanahauissa hakutuloksiin liitetään hakijan käyttämät avainsanat. Hän suosittelee myös, että kontrolloituja sanastoja tulee käyttää jatkossakin. (Hamburger, 2004, 91.)

Gracy (2014) mukaan käyttäjien auttaminen kontekstin ymmärtämisessä on tärkeää ja hän esittää tapoja, joilla sen voi tehdä. Käyttäjille tulee tarjota hakuvälineissä hänen mukaansa bibliografista tai historiallista tietoa asiakirjojen luoista, tietoa siitä kuinka hyvin aineisto dokumentoi luojan toimintaa tai tehtäviä sekä mihin asiakirjoja on käytetty ja millaisissa prosesseissa aineisto on luotu ja järjestetty. Hän toivoo käyttäjille tarjottavan tietoa myös aineiston säilytyksestä ja säilyttäjästä sen elinkaaren aikana. (Gracy, 2014, 7.)

Gracy (2014) suosittelee *linkitettyä dataa* (ks. 3.2) käytettäväksi hakuvälineissä. Gracy hakuvälineitutkimus osoitti, että luonnollisen kielen hakuväline (OpenCalais) tulkitse usein väärin syötteet esimerkiksi fraasihaut henkilön nimeksi. (Gracy, 2014, 26.) Linkitetyn datan kohdalla puhutaan usein niin sanotuista hakupisteistä, englanniksi access point. Tutkimustensa perusteella hakuvälineissä tulisi tarjota pääsy aineistoon esimerkiksi henkilön nimen, maantieteellisen paikan, tapahtuman ja ajan perusteella sekä näiden yhdistämisellä. Paikan nimistä tulisi tarjota kaikki historiallisetkin nimet. (Gracy, 2014)

3.3. Linkitetty data ja ontologiat haun tukena

Linkitettyllä datalla tarkoitetaan erilaisista lähteistä koottujen tietojen linkittämistä yhteen. Teknisenä vaatimuksena linkittämiselle on datan tallentaminen koneluettavaan muotoon ja sen täsmällisesti määrittelemisen. (Bizer, Heath, & Berners-Lee, 2009, 2.) Linkitetty data tarjoaa tutkijoiden mukaan ratkaisun aiheenmukaiseen hakuun auttaen hakijoita löytämään useampia aineistoja aiheesta. Linkitettyllä datalla voidaan saavuttaa semanttisesti rikkaita palveluja ja yhdistää useita kulttuuriperintöaineistoja. (Lindquist et al., 2013, 916.) Linkitetyn datan hyötyjä ovat Lindquistin ja muiden (2013) mukaan yhteentoimivuuden saavuttaminen verkossa eri lähteissä sijaitsevien alkuperäislähteiden kesken ja aineistojen rikastaminen ulkopuolisten lähteiden avulla. Lisäksi linkitetyn datan avulla voidaan tehdä älykkäitä palveluita, esimerkkinä semanttinen hakupalvelu sekä erilaisia suosituksia sisältöön liittyen. Data voidaan myös esittää monin eri tavoin. (Lindquist et al., 2013, 916-917.) Myös Amanda Hill (2004) on tutkimustensa perusteella suositellut eri järjestelmien välistä tiedonvaihtoa tutkijoiden tiedonhaun helpottamiseksi (Hill, 2004, 146).

Semanttinen tiedonhaku voidaan toteuttaa *ontologioiden* avulla. Ontologia voidaan määritellä jonkin ilmiön malliksi, jossa ilmiöön liittyvät käsitteet esitetään. Käsitteiden välille muodostetaan ontologiassa suhdeverkosto. Ontologioiden tarkoituksena on tavoittaa käsitteiden semanttinen taso koneluettavassa muodossa. (Jain & Singh, 2013, 65.) Nykyri ja Palonen (2014, 3-4) haastattelivat kuvailuammattilaisia, jotta ontologioiden kehittämisen sidosryhmien näkemykset tulevat huomioduksi. Ontologioiden sisältämiä täsmällisiä käsitteitä voidaan käyttää tiedonhaussa. (Nykyri & Palonen, 2014, 15). Haastateltujen näkemysten mukaan ontologiat ovat hyödyllisiä monin tavoin. Ontologioiden avulla haut voidaan automaattisesti täydentää hakusanoihin liittyvillä käsitteillä. Ontologiat myös yksiselitteistävät kuvailussa ja tiedonhaussa käytettyä sanastoa ja niiden monikielisyys koetaan eduksi. Joidenkin haastateltujen mukaan niitä voidaan tulevaisuudessa käyttää myös kokotekstihaussa. (Nykyri & Palonen, 2014 26-27.)

Informaation löydettävyyttä voidaan edistää myös käyttäjien luoman metadatan avulla. Sosiaalisten verkkotyöskentelytyökalujen lisääminen voi myös lisätä sivuston sisältöjen käyttöä. Tällainen sosiaalinen työkalu on esimerkiksi mahdollisuus lisätä niin sanottuja nimilappuja tai tageja sivuston sisältöihin. Digitaalisen kirjaston sisältöjen käyttäjistä etenkin asiantuntijat ja tutkijat voisivat hyötyä tästä. (Parandjuk, 2010, 129.)

4. TUTKIMUSASETELMA

Tässä luvussa kerrotaan tutkielmassa käytetystä tutkimusmenetelmästä. Luvussa kerrotaan myös mihin tutkimuskysymyksiin tutkielmassa pyritään vastaamaan ja miten niitä on rajattu. Alaluvussa 4.2 kerrotaan tutkielma-aineiston keräämisestä ja sen rajaamisesta. Tutkielma-aineiston analyysistä kerrotaan alaluvussa 4.3.

4.1. Tutkimuskysymykset

Tutkielmassa tarkastelen arkistokuvailumalleja arkistoaineiston haettavuuden ja löydettävyyden näkökulmasta. Arkistoaineiston käyttäjistä tehtyjen tutkimusten perusteella muodostan listan hakukriteereistä, joita käyttäjät tarvitsevat löytääkseen tarvitsemansa aineiston tai informaation. Tiedonhakututkimuksista saatua tietoa vertaan arkistokuvailussa käytettyyn nykyiseen kuvailumalliin ja kehitteillä olevaan uuteen arkistokuvailumalliin eli AHAA-käsitelmämalliin. Vertailussa mukana on myös nykyisen kuvailumallin mukainen JHS 143 -säännösten metatietomalli ja aineiston sähköisen säilyttämisen ohjeeksi tehty SÄHKE2-ohjeen metatietomalli sekä joitakin AHAA-kuvailusääntöjä.

Tutkimuskysymykset määritellään seuraavasti:

TK1: Mitkä ovat arkiston asiakkaiden kannalta keskeisiä hakukriteereitä tiedontarpeisiin tai hakemiseen kohdistuvan tutkimuksen perusteella?

TK2: Kuinka laajasti arkistoaineiston nykyinen kuvailumalli huomioi arkiston asiakkaiden hakukriteerit?

TK3: Kuinka laajasti AHAA-kuvailumalli huomioi arkiston asiakkaiden hakukriteerit?

TK4: Mitkä ovat olennaisimmat erot kuvailumallien välillä?

Tässä tutkielmassa ei käsitellä hakujärjestelmien käyttöliittymiä ja niiden muita ominaisuuksia. Oletuksena pidetään, että hakujärjestelmät ja niiden käyttöliittymät tarjoavat kaikki ne hakumahdollisuudet, jotka kuvailumallit ja kuvailun metatietomallit mahdollistavat. Pakolliseksi merkityt aineistosta tallennettavat tiedot siis mahdollistavat aina oletuksena hakemisen niiden perusteella. Ei-pakollisiksi eli vapaaehtoisiksi merkityt tallennettavat metatiedot vähentävät aineiston hakemisen mahdollisuuksia merkittävästi, koska vapaaehtoisiksi merkityt tiedot ei usein tallenneta (Kettunen & Henttonen, 2010).

4.2. Hakukriteerien kerääminen ja tutkielmaan valitut kuvailumallit

Hakukriteerien keräämiseksi etsin uusimpia tutkimustuloksia liittyen tiedonhaun muutoksiin ja arkistoaineistojen käyttäjiin. Tiedonhaun muuttuminen internetin hakukoneiden yleistymisen myötä on tutkimusten mukaan vaikuttanut myös arkistoaineistoja käyttävien hakutapoihin ja hakuprosesseihin. Arkistoaineiston käyttäjien näkökulma arkistokuvailuun ja sen hakuun on melko uusi tutkimusaihe. Suoraan arkistoaineiston hakijoiden käyttämiin hakutermeihin ja kriteereihin kohdistuvia tutkimuksia on tehty rajallinen määrä, mutta esimerkiksi Elenan ym. tekemä tutkimus (2011) kohdistui niihin. Edellä mainituista syistä olen ottanut tähän tutkielmaan mukaan uusimmat löytämäni tutkimukset arkistoaineiston käyttäjistä.

Kulttuuriperintöaineistojen kuvailu on uudistumassa osittain teknisen kehityksen seurauksena ja osittain paradigman muutoksen seurauksena (Henttonen & Kilkki, 2015, 8-10). Tästä syystä tutkielman tarkastelun kohteiksi on valittu nykyinen vuodelta 1997 oleva kuvailumalli (ks. luku 2.1) sekä siihen perustuva julkisille organisaatioille tehty JHS 143 -suosituksen metatietomalli (ks. luku 2.1.1) ja sähköisten asiakirjojen hallintaan ja siirtoon tehty SÄHKE2 -suosituksen metatietomalli (ks. luku 2.1.2). Julkisten organisaatioiden asiakirjallisen tiedon kuvailulle on tarkemmat määräykset ja säännöt kuin yksityishenkilöiden tai -yritysten arkistoitavalle tiedolle. Uuden paradigman mukainen arkistokuvailumalli on luonnosvaiheessa, eikä siitä ole johdettu vielä yhtenäistä kuvailusäännöstöä. Uudenlaisella kuvailumallilla on tarkoitus parantaa arkistoaineistojen kuvailumetatietoa myös aineiston käyttäjiä hyödyttävällä tavalla. Tämän vuoksi myös uusi AHAA-kuvailumalli ja joitakin (keskeneräisiä) AHAA-kuvailusääntöjä on tämän tutkielman tarkastelun kohteena.

4.3. Tutkielma-aineiston analyysi

Alla on tiivistetty lista kirjallisuudesta ja tutkimuksesta löydetystä hakijan tiedontarpeisiin liittyvistä hakukriteereistä. Tutkimuksista, joista hakukriteerit on poimittu, on kerrottu laajemmin luvussa 3. Tutkimuksissa on mainittu myös erilaisia kuvailutapoja ja niiden vaikutuksia hakumahdollisuuksiin. Julkaistujen tutkimusten perusteella tämän tutkielman tarkasteltaviksi hakukriteereiksi nousivat:

- aihe (J. Dooley, 1990 Schaffner, 2009, 6 mukaan; Hamburger, 2008, 84; Elena et al., 2011, 29 & 33)
- paikannimi tai sijainti (Duff & Johnson, 2003, 84; Gracy, 2014, 10; Schaffner, 2009, 7; Elena et al., 2010, 33)

- eri muodot, esimerkiksi paikan aikaisemmat nimet (Elena et al., 2010, 32)
 - aikaisemmat rajat (Duff & Johnson, 2003, 94)
- henkilön nimi (Hamburger, 2004, 84; Schaffner, 2009, 7; Gracy, 2014, 10; Duff & Johnson, 2003, 94; Elena et al., 2010, 32)
 - eri muodot
 - aikaisemmat nimet
- organisaatio (Elena et al., 2010, 32)
- ajankohta (Duff & Johnson, 2003, 87; Elena et al., 2010, 29)
 - dokumentin laatimisajankohta (Georgas, 2013, 512)
 - sisältöön liittyvä (Georgas, 2013, 512; Elena et al., 2010, 29)
- tapahtuma (ajallinen ilmiö) (Gracy, 2014, 18; Duff & Johnson, 2003, 94)
 - mikrotason tapahtuma
 - makrotason tapahtuma, voi koostua monista pienemmistä erillisistä tapahtumista (Gracy, 2014, 18)
- tieto asiakirjatyypin sisällöistä (Duff & Johnson, 2003, 91)
- tieto asiakirjatyypin luoista (Duff & Johnson, 2003, 91)
- aineiston formaatti (Georgas, 2014, 511-512)
- kontekstualisointi (Lindquist, Dulock, Törnroos, Hyvönen, & Mäkelä, 2013, 916; Lindquist & Long, 2011; Gracy, 2014, 7; Elena et al., 2010, 29 & 32)
 - eri tietojen yhdistäminen haettuun hakutuloksissa (Gracy, 2014, 7.)
 - esimerkiksi organisaatioiden ja instituutioiden historia ja nimet (Elena et al., 2010, 29)
 - käsitteiden synonyymit ja johdannaiset (Elena et al., 2010, 32)
 - historiallisten tapahtumien konteksti aineistoon (Lindquist & Long, 2011, 233; Duff & Johnson, 2003, 83-84; Elena et al., 2011, 33-34)
 - esim. jokin epidemia tai muu suureen osaan väestöstä vaikuttanut tapahtuma
 - aiheeseen liittyvät dokumentit (Elena et al., 2011, 33-34)
 - luojaan liittyvät tiedot (Elena et al., 2011, 33-34)
 - luontipäivään ja sisällön ajankohtaan liittyvät tiedot (Elena et al., 2011, 33-34)
 - sisältöä kuvaaviin termeihin liittyvät tiedot (Elena et al., 2011, 33-34)

Kuvailutavoista kirjoitettu on osittain ristiriitaista keskenään. Nykyisen tietämyksen mukaan kontrolloidut sanastot esimerkiksi ontologioiden avulla luotavana semanttisena verkkona yhdenmukaistavat kuvailua ja niiden avulla käsitteiden merkitykset ovat myös koneluettavassa

yksiselitteisessä muodossa. Kuitenkin kuvailutavasta riippumatta hakukonejärjestelmän ominaisuuksilla voi vaikuttaa haun tuloksiin suuresti.

Kuvailutavoista tutkimuksissa mainittiin seuraavia seikkoja:

- johdonmukaisuus (Elena et al., 2010, 32)
- kontrolloitu sanasto eli dokumentaatiokieli (Hamburger, 2004, 91; Lindquist et al., 2013, 916-917)
- luonnollinen kieli (Hamburger, 2004, 91)
- avainsanoitus (Hamburger, 2004, 84; Schaffner, 2009, 7)
 - eri järjestelmissä ja eri tutkimuksissa tarkoitetaan eri asioita (Lindquist et al., 2013, 915)
 - määritelmänä esim. ”*dokumentin sisältöä kuvaava termi*” (Alaterä & Halttunen, 2002,69)

Tässä tutkielmassa käsitellyt hakututkimukset ja -kirjallisuus jättävät hakukriteereihin puutteita, koska tutkimukset ovat keskittyneet vain tiettyihin hakemisen osa-alueisiin. Hakukriteereitä on sen vuoksi täydennetty esimerkiksi asiakirjojen perusominaisuuksilla kuten asiakirjan käsittelemällä luokitukseen perustuvalla asialla ja joillakin asiakirjan yhteyksillä muihin aineistoihin.

AHAA-käsitelmän tarkoitus on kuvata ylimmällä tasolla arkistokuvailun tietovarantoentiteetit ja kontekstientiteetit sekä niiden väliset suhteet (Henttonen & Kilkki, 2013, 19). Nykyinen kuvailumalli on määritellyt kuvailuelementit (ks. taulukko 1) konkreettisemmalla tasolla. Tästä seuraa se, etteivät kuvailumallit ole täydellisesti vertailukelpoisia. Lisäksi mallit eivät ota kantaa kuvailutapaan.

Tutkielmassa vertaillaan hakukriteerejä myös julkiselle hallinnolle tehdylle JHS 143 -suosituksen metatietomalliin ja keskeneräisiin uusiin kuvailusääntöihin. Tulevat AHAA-käsitelmän mukaiset kuvailusäännöt ja metatietomallit ovat moniulotteisia, joten niillä pystytään kuvaamaan myös monenvälisiä kontekstien suhdeverkostoja. Tässä tutkielmassa vertailuun käytettävät JHS 143 -suosituksen metatietomalli ja SÄHKE2-metatietomalli ovat yksiulotteisia. (Henttonen & Kilkki, 2013, 24.)

5. TULOKSET

Luvussa esitellään hakututkimusten perusteella tehdyn listan (edellinen luku) avulla laadittu taulukko, jota on täydennetty hakututkimusten osittaisen vaillinaisuuden vuoksi. Taulukon hakukriteereitä verrataan nykyiseen kuvailumalliin, JHS 143 -suosituksen metatietomalliin, SÄHKE2-määräyksen metatietomalliin sekä osittain keskeneräiseen AHAA-kuvailumalliin ja siitä johdettuihin keskeneräisiin kuvailusääntöjen metatietoihin.

5.1. Arkistoaineiston käyttäjien hakukriteerit

Taulukossa 5 (Hakukriteerit ja kuvailutavat) on lueteltu tiedonhakijoiden tiedonhakutarpeista kumpuavat hakukriteerit ja niiden kuvailutavat. Taulukossa käsitellään samoissa kohdissa sekä yksittäisen asiakirjan kuvailutietoja että suurempiin kokonaisuuksiin liittyviä kuvailutietoja. Kuvailun ja aineistojen hierarkkisuuuteen ei ole muulla tavalla otettu kantaa.

Hakutarpeiden perusteella taulukossa 5 on otettu kantaa myös kuvailutapaan. Tutkimusten perusteella hakijat haluavat hakea sekä vapaasanahauulla, avainsanahauulla että kontrolloitua sanastoa käyttämällä (Alaterä & Halttunen, 2002; Hamburger, 2004; Schaffner, 2009). Näitä vastaavat kuvailutavat ovat vapaa teksti, nimetty tietoalkio ja kontrolloitu tietoalkio. Kuvailumalleissa ja kuvailusääntöjen metatietomalleissa ei kuitenkaan ole aina määritetty suoraan metatietojen kuvailutapaa tässä merkityksessä tai siihen on annettu vaihtoehtoisia tapoja. Näin on tehty SÄHKE2-metatietomallissa ja JHS 143 -suosituksessa. AHAA-käsitemalli on määritellyt useimmat entiteetit tallennettavaksi kontrolloituna tietoalkiona.

Taulukko 5: Hakukriteerit ja kuvailutavat.

Hakijan tiedontarve	Kuvailutapa
1. Mitä asiakirja/aineisto käsittelee?	
aihe	vapaa teksti
	nimetty tietoalkio
	kontrolloitu tietoalkio
henkilö, toimija	vapaa teksti
	nimetty tietoalkio
	kontrolloitu tietoalkio
paikka	vapaa teksti
	nimetty tietoalkio
	kontrolloitu tietoalkio
tapahtuma/ilmiö	vapaa teksti
	nimetty tietoalkio
	kontrolloitu tietoalkio
ajankohta	vapaa teksti
	nimetty tietoalkio
	kontrolloitu tietoalkio
2. Asiakirjan/aineiston ominaisuudet (luettelointitiedot)	
asiakirjatyyppe	kontrolloitu tietoalkio
identifiointitunnus	kontrolloitu tietoalkio
nimeke/nimi	kontrolloitu tietoalkio
laatimispäivämäärä	kontrolloitu tietoalkio
asiatunnus	kontrolloitu tietoalkio
asian nimi	kontrolloitu tietoalkio
formaatti	kontrolloitu tietoalkio
kieli	kontrolloitu tietoalkio
3. Toimijatiedot	
arkistonmuodostaja(t)	kontrolloitu tietoalkio
arkistonmuodostajan tehtävät	vapaa teksti
	nimetty tietoalkio
	kontrolloitu tietoalkio
henkilön nimi	vapaa teksti
	nimetty tietoalkio
	kontrolloitu tietoalkio
organisaatio	vapaa teksti
	nimetty tietoalkio
	kontrolloitu tietoalkio
henkilön/organisaation rooli	kontrolloitu tietoalkio
4. Paikkatiedot	
luontipaikka	kontrolloitu tietoalkio
muokkauspaikka	kontrolloitu tietoalkio
5. Asiakirjatyypitiedot	
nimi	kontrolloitu tietoalkio

sisältö	kontrolloitu tietoalkio
luoja	kontrolloitu tietoalkio
6. Asiakirjan yhteydet	
asiakirjatyypin tietoihin	
paikkoihin	
paikkojen aikaisemmat nimet ja rajat	
luontiajankohdan tapahtumiin	
sisällöllisen ajankohdan tapahtumiin	
osa- ja kokonaissuhteet	
henkilöihin	
henkilöiden aikaisemmat nimet ja eri muodot	
organisaatioihin	
muihin asiakirjoihin	
muihin aineistoihin	
samaa aihetta /asiaa käsitteleviin asiakirjoihin	
julkaisuihin	

Hakukriteereitä verrataan seuraavissa luvuissa kuvailumallien kuvailuelementteihin. Kuvailumallien antama mahdollisuus tallentaa tietoja kuvailuelementteinä ei käytännössä tarkoita sitä, että tallennetun metatiedon avulla voisi hakea aineistoa. Tähän vaikuttavat metatiedon tallentamisen pakollisuus tai vapaaehtoisuus sekä itse hakujärjestelmän ja sen käyttöliittymän hakuominaisuudet. Vapaaehtoiset tiedot on merkitty vertailutaulukoissa merkillä (x) ja pakolliset kuvailutiedot merkillä x. Näiden perään on merkitty, mikä kuvailumallin tai metatietomallin elementti hakukriteerin kohdalla tulkittu vastaavaksi.

5.2. Nykyisen kuvailumallin, JHS 143 -metatietomallin ja SÄHKE2-metatietomallin vertailu hakukriteereihin

Taulukossa 6 (Hakukriteerien ja nykyisen kuvailumallin ja metatietomallien vertailu) on käyty läpi kuvailumallien metatietoelementit vertaillen niitä tutkimuksista löydettyihin hakukriteereihin. Kuvailutapa ei aina ilmene kuvailumallista tai metatietomalleista, joten osa kohdista on merkitty mahdollisiksi toteuttaa kuvailumallin perusteella.

Taulukko 6: Hakukriteerien ja nykyisen kuvailumallin sekä metatietomallien vertailu

Hakijan tiedontarve	Kuvailutapa	Nykyinen kuvailumalli	JHS 143 -metatieto-malli	SÄHKE2-metatieto-malli
1. Mitä asiakirja/aineisto käsittelee?				
aihe	vapaa teksti	x (tietosisältö)	(x) (kuvaus)	(x) (kuvaus)
	nimetty tietoalkio	x (hakusanat)	(x) (aihe)	(x) (aihe)
	kontrolloitu tietoalkio		(x) (aihe)	(x) (aihe)
henkilö, toimija	vapaa teksti	x (historia)	x (kuvaus)	(x) (kuvaus)
	nimetty tietoalkio			
	kontrolloitu tietoalkio			
paikka	vapaa teksti	(x) (maantiet. toiminta-alue)	(x) (kuvaus)	(x) (kuvaus)
	nimetty tietoalkio	mahdollinen	(x) (kattavuus)	(x) (aihe)
	kontrolloitu tietoalkio	mahdollinen	(x) (kattavuus)	(x) (aihe)
tapahtuma/ilmiö	vapaa teksti			(x) (kuvaus)
	nimetty tietoalkio			
	kontrolloitu tietoalkio			
ajankohta	vapaa teksti	x (tietosisältö)	(x) (kattavuus)	(x) (kuvaus)
	nimetty tietoalkio		(x) (kattavuus)	(x) (aihe, aikamääre)
	kontrolloitu tietoalkio		(x) (kattavuus)	(x) (aihe)
2. Asiakirjan/aineiston ominaisuudet (luettelointitiedot)				
asiakirjatyyppi	kontrolloitu tietoalkio	x (nimet)	(x) (ak:n tyyppi)	x (toimenp. tyyppi)
identifiointitunnus	kontrolloitu tietoalkio	(x) (tunnus)	x (ident. tunnus)	x (ident. tunnus)
nimeke/nimi	kontrolloitu tietoalkio	x (nimet)	x (nimeke)	x (nimeke)
laatimispäivämäärä	kontrolloitu/nimetty tietoalkio	x (aikamääreet)	x (toimenpiteen ajankohta)	x (aikamääre)
asiatunnus	kontrolloitu tietoalkio		x (tehtävä)	x (6.3.4)
asian nimi	kontrolloitu tietoalkio	(x) (tehtävät)	x (nimeke)	x (6.3.5)
formaatti	kontrolloitu tietoalkio	(x) (fyysinen muoto)	x (formaatti)	x (formaatti)
kieli	kontrolloitu tietoalkio	x (asiakirjakieli)	(x)(kieli)	x (kieli)

3. Toimijatiedot				
arkistonmuodosta- ja(t)	kontrolloitu tietoalkio	x (nimet)	x (toimija, yhteisön nimi)	x (toimija)
arkistonmuodosta- jan tehtävät	vapaa teksti	(x) (tehtävät)		
	nimetty tietoalkio	mahdollinen	x (tehtävä)	
	kontrolloitu tietoalkio	mahdollinen	x (tehtävä)	x (tehtävä)
henkilön nimi	vapaa teksti	x (nimet)		
	nimetty tietoalkio	mahdollinen	x (toimija)	
	kontrolloitu tietoalkio	mahdollinen	x (toimija)	x (toimijan henkilö)
organisaatio	vapaa teksti	x (nimet)		
	nimetty tietoalkio	mahdollinen	x (toimijan tunniste)	
	kontrolloitu tietoalkio	mahdollinen	x (toimijan tunniste)	(x) (toimijan yhteisö)
henkilön/organisaat- ion rooli	kontrolloitu tietoalkio		x (toimijan rooli)	(x) (toimijan rooli)
4. Paikkatiedot				
luontipaikka	kontrolloitu tietoalkio	(x) (maantiet. toiminta-alue)		
säilytyspaikka	kontrolloitu tietoalkio	(x)(omistushis- toria, luovutustiedot)	x (sijaintipaikka)	(x) (sijaintipaikka)
5. Asiakirjatyypitiedot				
nimi	kontrolloitu tietoalkio	x (nimet)	x (ak:n tyyppi)	x (tyyppi)
sisältö	kontrolloitu tietoalkio			x (tyyppi)
luoja	kontrolloitu tietoalkio			
6. Asiakirjan yhteydet				
asiakirjatyypin tietoihin				
paikkoihin				
paikkojen aikaisemmat nimet ja rajat				
luontiajankoh- dan tapahtumiin				
sisällöllisen ajankohdan tapahtumiin				
osa- ja kokonaissuhteet			x (tehtävän rekisterinro)	x (tehtävä- luokitus)
henkilöihin			(x) (toimija: muu tekijä)	
henkilöiden aikaisemmat				

nimet ja eri muodot				
organisaatioihin		x (edeltäjät/seuraajat)	(x) (toimija: muu tekijä)	
muihin asiakirjoihin		(x) (liittyvät aineistot)	(x) (lähde, suhde)	(x) (vain eri versioon)
muihin aineistoihin		(x) (muualla säilytettävä aineisto)	(x) (suhde)	(x) (viittaus kohteeseen)
samaa aihetta /asiaa käsitteleviin asiakirjoihin		(x) (liittyvät aineistot)		
julkaisuihin		(x) (julkaisut/tutkimukset)	(x) (suhde)	

Aineistolla voi olla nykyisen kuvailumallin ja metatietomallien mukaan vain yksi arkistonmuodostaja eli vanhan paradigman mukaisesti yksi proveniensi. Sisältötietojen alueen hakusanat on tulkittu avainsanoiksi ja tietosisältö sekä historia tai elämäkerta luonnollisen kielen kuvailutavaksi. Tietosisältö- ja historia tai elämäkerta -elementteihin käytännössä sisällytettävä tieto todennäköisesti vaihtelee suuresti aineistoittain. Taulukossa 6 nämä elementit on tulkittu siten, että niihin voi vapaana tekstinä kuvailla aiheen lisäksi toimijan eli organisaation tai henkilön ja sen ajankohdan, mitä aineisto käsittelee.

Liittyviä organisaatioita nykyisessä kuvailumallissa voivat olla edeltäjät ja seuraajat, mikä tarkoittaa siis arkistonmuodostajaa (Arkistolaitos, 1997b). Taulukossa 6 on aineiston yhteyksien kohtaan merkitty mahdollisuus viitata muihin aineistoihin elementillä liittyvät aineistot, mikä ei ole pakollinen. Kaikki viittaussuhteet ovat käytännössä nimettyjä tietoalkioita tai vapaata tekstiä, mikä ei välttämättä auta järjestelmien kautta tehtävässä tiedonhaussa.

Aihe tallennetaan JHS 143 -metatietomallin mukaan asiasanoina tai luokittelujärjestelmää (AMSin tehtävä- tai asiaryhmittely) käyttäen. Aihe on pakollinen vain asiakirjan julkaisemisessa, ei asiakirjahallinnolle. (JHS 143, 2006, 10.) Kattavuus merkitsee JHS 143:n metatiedoissa joko hallinnon alaa tai maantieteellistä kattavuutta. Se ei ole pakollinen metatieto ja se voidaan tallentaa nimettyyn tietoalkioon tai merkitä eri lähteistä saatavilla koodeilla. Sillä voidaan merkitä myös ajallista kattavuutta. (JHS 143, 2006, 21.) Nämä on merkitty taulukkoon 6 asiakirjan tai aineiston käsittelemäksi paikaksi tai ajaksi.

Taulukkoon 6 merkitty yhteys muihin asiakirjoihin on mahdollinen JHS 143:n lähde- ja suhde-metatiedon kautta. Se suositellaan merkittäväksi URI-tunnuksella. Suhde-metatiedolla voidaan

merkitä muun tyyppiset suhteet ja samaan asiaan liittyvät asiakirjat merkitään rekisteritunnuksen avulla. Suhde on pakollinen metatieto, jos sillä on merkitystä asiakirjan elinkaaren kannalta. Suhde-metatiedon kohdalla määritellään sen kuvaavan kuitenkin lähinnä samaan aktiin liittyvien asiakirjojen suhdetta, ei samaan aiheeseen liittyvää suhdetta. (JHS 143, 2006, 29 & 34.) Tämän vuoksi taulukkoon ei ole merkitty sen liittävän samaa aihetta käsittelevien asiakirjojen suhdetta. JHS 143 -suosituksen sisällönkuvailun nimeke on tulkittu asiakirjan käsittelemäksi asiaksi, koska se on tiivistetty kuvaus asiakirjan sisällöstä tai tarkoituksesta (JHS 143, 2006, 30).

SÄHKE2 -metatietomallissa aihe on määritelty kirjattavaksi asiasanoina tai kontrolloidun sanaston mukaisesti, mutta se on vapaaehtoinen. Metatietoihin sisältyy myös metatieto kuvaus, johon voi vapaaehtoisena kuvailla sisällön luonnollisella kielellä. Nämä kaikki on otettu huomioon kohdassa asiakirjan käsittelemissä tiedoissa aiheen kohdalla. SÄHKE2:n alueellinen kattavuus voidaan merkitä joko luonnollisena kuvailuna tai erilaisten standardien koodien avulla. Ne ovat siis keskenään vaihtoehtoisia ja metatieto on valinnainen. Metatietomallin mukaan viittaaminen muihin asiakirjoihin on mahdollista, mutta tällä tarkoitetaan lähinnä eli kieliversioon viittaamista. Tehtävuokituksen tunnus (6.3.4) ja nimeke (6.3.5.) on merkitty taulukossa 6 asiatusunukseksi ja asian nimeksi. (Arkistolaitos, 2009a, 6-24.)

5.3. AHAA-kuvailumallin ja -kuvailusääntöjen vertailu hakukriteereihin

Taulukossa 7 hakukriteereitä verrataan luonnosvaiheessa olevaan AHAA-arkistokuvailun käsittemalliin sekä myös keskeneräisiin AHAA-kuvailusääntöihin. AHAA-kuvailusäännöistä ei ollut vielä saatavilla yhtenäistä dokumenttia, joten vertailu perustuu moneen eri työdokumenttiin. AHAA-käsittemalli ei ota kantaa entiteetteihin sisältyvien metatietojen pakollisuuteen. AHAA-kuvailusääntöjen keskeneräisyydestä johtuen kaikkia hakukriteerien mukaisia kohtia ei ollut mahdollista tässä vaiheessa arvioida. AHAA-kuvailusääntöjenkään työdokumentit eivät ota kantaa metatietojen pakollisuuteen, joten kaikki niistä löytyneet on merkitty merkillä x (perässä on kuvailumetatiedon nimi).

AHAA-kuvailumallin versio 0.2 (Henttonen & Kilkki, 2015) nimeää seuraavat entiteettien tunnukset (mainittu taulukossa 7):

- E1 Toiminta
 - E2 Tehtävä
 - E3 Elinkaarihallinta
- E4 Valtuutus
- E5 Toimija
- E6 Paikka
- E7 Tapahtuma
- E8 Aineisto
 - E9 Manifestaatio
- E10 Aihe

Taulukko 7: Hakukriteerien vertailu AHAA-kuvailumalliin ja -kuvailusääntöihin.

Hakijan tiedontarve	Kuvailutapa	AHAA-käsittemalli	AHAA-kuvailusäännöt
1. Mitä asiakirja/aineisto käsittelee?			
aihe	vapaa teksti		x (tietosisältö)
	nimetty tietoalkio		
	kontrolloitu tietoalkio	x (E10)	x (asiasanat)
henkilö, toimija	vapaa teksti		(x) (tietosisältö)
	nimetty tietoalkio		
	kontrolloitu tietoalkio	x (E5)	x (toimija)
paikka	vapaa teksti		x (esim. toimijan tiedot)
	nimetty tietoalkio		
	kontrolloitu tietoalkio	x (E6)	x
tapahtuma/ilmiö	vapaa teksti		x (tietosisältö)
	nimetty tietoalkio		
	kontrolloitu tietoalkio	x (E7)	x (liitetään ajallisen ilmiön tietoihin)
ajankohta	vapaa teksti		
	nimetty tietoalkio	x, mahd.	x, mahd.
	kontrolloitu tietoalkio	x, (E1, E7, E8, E9, E10)	x
2. Asiakirjan/aineiston ominaisuudet (luettelointitiedot)			
asiakirjatyyppe	kontrolloitu tietoalkio		
identifiointitunnus	kontrolloitu tietoalkio	x (E8, E9)	(x, viitekoodit)
nimeke/nimi		x (E8, E9)	x
laatimispäivämäärä	kontrolloitu tietoalkio	x (E1)	x
asiatunnus	kontrolloitu tietoalkio	x (E2)	ei mainintaa
asian nimi	kontrolloitu tietoalkio	x (E2)	ei mainintaa
formaatti	kontrolloitu tietoalkio	x (E9)	x (aineiston fyysinen luonnehdinta ja tekniset ominaisuudet)
kieli	kontrolloitu tietoalkio	x (E9)	
3. Toimijatiedot			
arkistonmuodostaja(t)	kontrolloitu tietoalkio	x (E5)	x (toimijat)
arkistonmuodostajien tehtävät	vapaa teksti		x (kattavuus ja sisältö)
	nimetty tietoalkio		
	kontrolloitu tietoalkio	x (E2)	
henkilön nimi	vapaa teksti		
	nimetty tietoalkio		

	kontrolloitu tietoalkio	x (E5)	x (toimija)
organisaatio	vapaa teksti		
	nimetty tietoalkio		
	kontrolloitu tietoalkio	x (E5)	x (toimija)
henkilön/organisaation rooli	kontrolloitu tietoalkio	x (E5)	x (liitetään toimijaan)
4. Paikkatiedot			
luontipaikka	kontrolloitu tietoalkio	x (E6)	
säilytyspaikka	kontrolloitu tietoalkio	x (E6)	
5. Asiakirjatyypitiedot			
nimi	kontrolloitu tietoalkio	mahdollinen (E8)	
sisältö	kontrolloitu tietoalkio	mahdollinen (E8)	
luoja	kontrolloitu tietoalkio	mahdollinen (E8)	
6. Asiakirjan/aineiston yhteydet			
asiakirjatyypin tietoihin		mahdollinen (E8)	ei mainintaa
paikkoihin		x (E6)	ei mainintaa
paikkojen aikaisemmat nimet ja rajat		x (E6)	ei mainintaa
luontiajankohdan tapahtumiin		mahdollinen	x (ajallisen ilmiön tiedot)
sisällöllisen ajankohdan tapahtumiin		mahdollinen	x (ajallisen ilmiön tiedot)
osa- ja kokonaissuhteet		x (E9)	x (yhteen kuuluvat kuvailuyksiköt)
henkilöihin		x (E5)	x (Ensisijainen nimi, tunnus)
henkilöiden aikaisemmat nimet ja eri muodot		x (E5)	x (vaihtoehtoinen nimi)
organisaatioihin		x (E5)	x
muihin asiakirjoihin		mahdollinen	x
muihin aineistoihin		mahdollinen	mahdollinen
samaa aihetta /asiaa käsitteleviin asiakirjoihin		mahdollinen	mahdollinen
arkistonmuodostajien tietoihin		x (Toimijan suhteet)	x
julkaisuihin		mahdollinen	x (julkaisuviitteet)

AHAA-käsitellessä määritelty entiteetti E10_Aihe (Subject) kattaa monenlaisia sisältöjä. Siihen sisältyy abstraktit käsitteet ja materiaaliset objektit eli muun muassa toimijat, tehtävät, toiminnot ja paikat. Lisäksi se käsitellään myös kontekstin representaatioksi. (Henttonen & Kilkki, 2015, 27-28.) Entiteetti E6_Paikka on määritelty historialliseksi tai nykyiseksi maantieteelliseksi tai hallinnolliseksi alueeksi ja se voi olla sekä fyysinen että virtuaalinen tila. Implementoinnissa se on ontologiana kuvattavaa auktoriteettitietoa. (Henttonen & Kilkki, 2015, 24-25.) Entiteetti E5_Toimija (Agent) voi olla yhteisö, suku, perhe sekä luonnollinen henkilö tai myös järjestelmä. Toimija on toteutuksessa auktorisoitua tietoa. (Henttonen & Kilkki, 2015, 23-24.)

Entiteettiin E8_Aineisto on tulkittu sisältyvän mahdollisena asiakirjatyypin tiedot taulukossa 7, koska se mallintaa tietovarantoa loogisena entiteettinä (Henttonen & Kilkki, 2015, 26). AHAA-käsitellessä ei mainitse asiakirjatyypin kuvailua erikseen. Kuvailusäännöissä ei myöskään ainakaan tähän mennessä ole mainintaa asiakirjatyypin tai aineistotyypin kuvailusta erillisenä. Asiakirjan tai aineiston yhteydet taulukossa 7 kuvattuihin tietoihin ovat kirjoittajan tulkinnan mukaan kaikki vähintään mahdollisia AHAA-käsitellessä puitteissa. Käsitellessä varsinaisesti kuvatut yhteydet ovat aineiston osa- ja kokonaissuhteet, yhteydet toimijoihin ja paikkoihin sekä niiden eri muotoihin auktorisoitujen tietojen avulla.

Työdokumentissa *Toiminta30052013* (2013) on määritelty sääntöjä toiminnan dokumentoimiseksi. Toiminnan konteksti ja toiminnan suhteet muuhun toimintaan tuovat selkeän parannuksen aiempiin kuvailusääntöihin. Niihin sisältyvät mm. metatiedot toiminnan ajallinen kesto ja historia, suhteen luokka, suhteen kuvaus sekä suhteen ajallinen kesto. Myös itse kuvailusta tallennettava metatieto on tarkemmalla tasolla kuin aikaisemmin. Toiminnan kontekstin kuvaus tehdään vapaamuotoisesti. (*Toiminta30052013*, 2013, 3-7.)

Tietosisältö ja asiasanat on määritelty työdokumentissa Kuvailusäännöt_Sisältö- ja rakennetietojen alueet (Lampela, 2013). Tietosisältö kuvaillaan vapaamuotoisesti ja se on entiteetin ekspressio ominaisuus, tämä on tulkittu taulukossa 7 aiheeksi, joka kuvaillaan vapaalla tekstillä. Asiasanat tulee dokumentin mukaan ottaa käytännössä ontologiasta, tämä on tulkittu taulukossa 7 aiheeksi kontrolloidun kuvailun kohdalle.

AHAA-kuvailusäännöissä ("Kuvailusäännöt-Tietovaranto") on määritelty myös aineistoon suoraan liittyvän ajallisen ilmiön liittäminen aineistoon. Toiminnan tiedot on tässä versiossa määritelty vapaaksi tekstiksi ja tulkittu taulukossa 7 asiaksi, jota aineisto tai asiakirja käsittelee.

5.4. Kuvailumallien vertailua

Nykyisessä kuvailumallissa kuvailun tekijälle ja mahdolliselle tietojärjestelmälle on jätetty enemmän valinnan varaa varsinaisen kuvailun toteutukseen. Tästä voi seurata helposti se, että eri aineistojen tai eri säilyttävien arkistojen tekemä kuvailu on keskenään erilaista. Tämä vähentää huomattavasti kuvailun johdonmukaisuutta. Kuvailun johdonmukaisuuden puute johtaa aineistojen hakemisen vaikeuksiin (Elena et al., 2010, 32).

Nykyinen kuvailumalli antaa mahdollisuuksia aineiston hakemiseen, mutta siinä enemmän vastuuta kuvailijan tulkinnoille ja kuvailijan antamille merkityksille eri käsitteistä. Esimerkiksi aineiston tai asiakirjan käsittelemään aiheeseen on mahdollista kuvailla vapaasti aineiston käsittelemiä henkilöitä, organisaatioita ja paikkoja. JHS 143 -suosituksen metatietomallin mukaisesti kuvailtuna on mahdollista kuvailla aineiston aihetta. Kuitenkin esimerkiksi aineiston käsittelemät paikka ja ajankohta on määritelty vapaaehtoisiksi. Kuvailuelementin vapaaehtoiseksi merkitsemisessä voi kuitenkin olla monia syitä. Suuri puute nykyisessä kuvailumallissa on erilaisten ajallisten ilmiöiden kuvailun mahdollisuuden puute erillisenä aineistosta. Kuvailumallissa ja metatietomalleissa on kuvaus-elementissä mahdollista tällaisten asioiden kuvailu. Se on kuitenkin kiinteä osa aineiston sisällön kuvausta, ei tapahtuman kuvaamista erillisenä ilmiönä. (Arkistolaitos, 2009a, 8.)

SÄHKE2:n metatietomallissa kaikki asiakirjan tai aineiston sisältöön liittyvät kuvailuelementit on merkitty vapaaehtoisiksi (Arkistolaitos, 2009a, 6 & 8). Käyttäjän kannalta tämä on ongelmallista, koska aineiston tiedollinen sisältö on heille olennaisinta. Toimenpiteen tyyppi on pakollinen, ja se kertoo myös asiakirjan oletettavasta sisällöstä jonkin verran asiakirjatyypin lisäksi. AHAA-kuvailusäännöissä ei ole toistaiseksi mainittu asiakirjatyypin määrittelyä.

AHAA-käsittemallissa on annettu käytännön ohjeita siitä, mitkä mallissa tehdyt entiteetit tulee tallentaa kontrolloituna tietona. Useimmiten tämä ilmaistaan siten, että entiteetin tiedot ovat auktorisoitua tietoa, joka kuvataan ontologiana. Tämä tuo selkeästi johdonmukaisuutta sekä kuvailujärjestelmien että kuvailun tekemiseen. Hakukriteeritaulukoissa tämä näkyy harvempina mahdollisina kuvailutapoina, vaikka se käytännössä tarkoittaa kuitenkin parempaa haettavuutta.

AHAA-kuvailumallissa on kuvattu entiteettien välisiä suhteita, nykyisessä kuvailumallissa nämä puuttuvat lähes täysin. Suhteet edesauttavat kontekstin luomista aineistolle. AHAA-kuvailumalli määrittää kuitenkin lähinnä yhden aineiston kuvailua, joten erillisten aineistojen väliset suhteet eivät ole määriteltyinä siinä hakukriteeritaulukossa kuvaamallani tavalla. Eri aineistojen väliset

suhteet voidaan kuitenkin auktorisoitujen tietojen kautta määrittää ja niitä on mahdollista käyttää hakemisen apuna järjestelmissä.

Kokonaisuutena tähän saakka tehdyt AHAA-kuvailusäännöt ovat vielä keskeneräiset ja vajaat, joten vertailua hakukriteereihin ei voida tässä tutkielmassa tehdä täysimääräisesti. Synonymian ja esimerkiksi erilaisten kirjoitusmuotojen ja aikaisempien nimien huomioon ottaminen on paremmin määritelty AHAA-käsitelmissä.

6. TULOSTEN TARKASTELUA

Tässä luvussa eritellään tuloksista tehtyjä havaintoja ja arvioidaan niiden merkitystä. Luvussa pohditaan myös kuvailumallien kehittämistarpeita haettavuuden näkökulmasta.

AHAA-kuvailumallissa ja -kuvailusäännöissä on selkeästi otettu huomioon tässä tutkielmassa näkökulmana oleva arkistoaineiston käyttäjien hakutarpeet. Nykyiset kuvailumallit ovat tämän tutkielman perusteella tehty enemmän aineiston hallintaa varten kuin käyttäjien hakutarpeita ajatellen. Muutos on seurausta sekä tekniikan kehityksestä että ajattelutavan muutoksesta.

AHAA-kuvailumalli osoittautui käyttäjien hakukriteerien näkökulmasta kattavammaksi kuin nykyinen vuodelta 1997 oleva kuvailumalli. Uuden kuvailumallin paremmuus tulee ilmi ennen kaikkea erillisinä tallennettavien tietojen välisten suhteiden lukumäärältään suurempana ja monimuotoisempana kuvailuna. AHAA-käsitemalli sopii myös paremmin erityyppisten aineistojen kuvailuun, nykyinen kuvailumalli on tarkoitettu suureksi osaksi asiakirjallisen aineiston kuvailuun.

Uuden käsitemallin mukaisessa kuvailussa metatiedot tallennetaan useimmiten ontologioissa käytettävillä auktoriteettitiedoilla. Hakujärjestelmä pystyy kuitenkin mahdollistamaan hakemisen myös luonnollisen kielen sanoilla ontologioista, joten hakemisen ei tarvitse rajoittua ontologian määrittelemiin sanoihin. Hakukone voi esimerkiksi ehdottaa ontologiassa käytettävää pääkäsitettä hakusanan tilalle tai hierarkiassa ylempänä tai alempana olevia käsitteitä.

Nykyisessä kuvailumallissa on jätetty määrittelemättä suureksi osaksi se, mihin haku voi kohdistua. Tämä johtuu osittain siitä, että kuvailumalli on tehty aineiston hallintaa varten, ei aineistojen käyttäjiä ajatellen. Sisältötietojen alueeseen kuuluva hakusanat-elementti (2.5) ei täytä nykyisen ajattelutavan mukaisia aineiston haettavuuden kriteereitä riittävästi.

Tapahtuma tai ilmiö aineistojen ulkopuolisena toimintana puuttuu nykyisistä kuvailu- ja metatietomalleista. Aineistojen suhde aikakautensa ilmiöihin olisi hyvä tuoda kuitenkin esille kontekstin lisäämiseksi. AHAA-kuvailusäännöissä otetaan kaikki aineistoon liittyvä toiminta hyvin huomioon, sillä kuvailuelementtien erillisyys toisistaan ja aineistosta sekä niiden linkittäminen toisiinsa parantavat haettavuutta.

AHAA-käsitemallia on mahdollista tulkita siten, että siinä otettu huomioon myös asiakirjatyypin, esimerkkinä päätöksen, kuvailu erillisenä aineistosta ekspression tasolla eli E8_Aineisto-elementissä (Henttonen & Kilkki, 2015, 26). Suoraa mainintaa tällaisesta ei kuitenkaan ole käsitemallissa, eikä myöskään tähän mennessä tehdyissä kuvailusäännöissä. Tämä olisi siis Duffin

ja Johnsonin näkemyksen (2003, 91) sekä tämän tutkielman perusteella parannettava asia AHAA-kuvailumallissa ja -säännöissä. Asiakirjatyyppejä on Duffin ja Johnsonin tutkimuksen perusteella merkittävä sukututkijoille (2003). Laajemman tutkimuksen tekijät eivät välttämättä hyödy asiakirjatyyppeiden kuvailusta.

Uusi kuvailusäännöstö on vielä hyvin keskeneräinen. Tämän tutkielman perusteella kuvailutavan muutos on kuitenkin menossa parempaan suuntaan. Haku- ja kuvailujärjestelmien kehittäminen uuden kuvailutavan mukaiseksi on tärkeää pitkällä tähtäimellä.

Samanlaista aiempaa tutkimusta ei löytynyt arkistokuvailumallien osalta. Kirjastokuvailumallin muutoksesta eli FRBR-mallin vaikutuksista aineiston kuvailutietojen näyttämiseen ja haettavuuteen on kirjoittanut Antonella Trombone (2014). Tromboneen mukaan FRBR tarjoaa entiteetti-relaatio-tietomallin, jonka mukaisesti kuvailtuna bibliografiset tiedot muodostavat navigoitavan rakenteen myös käyttöliittymään ja mahdollistavat paremmin internetin hakukoneiden haun toisin kuin nykyiset MARC-tiedostot esimerkkinä (Trombone, 2014, 26-28). Linkitettävän tiedon avulla voidaan esimerkiksi rikastaa hakutuloksissa näytettävää tietoa (Trombone, 2014, 24). Hakutuloksia on helpompi ryhmitellä teoksen mukaisesti, eikä tuloksissa esimerkiksi tarvitse näyttää erikseen jokaista kirjan painosta. Teokset voidaan siis ryhmitellä esimerkiksi ekspressioittain, mikä helpottaa teoksen hakijan työtä. (Trombone, 2014, 25-26.)

Arkistokuvailustandardien kehitys on vaikuttanut RDA:n kehitykseen (Nimer, 2010, 232). RDA:n vaikutuksia arkistoaineiston kuvailuun ja aineiston hakuun ei ole vielä voitu tutkia, mutta Nimer (2010) on pohtinut sen mahdollisia vaikutuksia arkistoille. Hän toteaa, että RDA-perusteisessa kuvailussa on mahdollisuuksia parempaan aineistohallintaan pienissä arkistoissa. RDA myös rikastaa arkistokuvailua ja mahdollistaa uusia tapoja linkittää esimerkiksi entiteettejä, internetissä sijaitsevia aineistoja ja toimijoita. (Nimer, 2010, 239-240.)

Odotetusti tässä tutkielmassa tehdyn tarkastelun ja tutkimustavan perusteella uusi arkistokuvailumalli vastaa käyttäjien hakutarpeisiin paremmin kuin vanha. Uusi kuvailumalli on myös odotetusti nykyisen ajattelutavan mukaisesti rakennettu ja vastaa melko pitkälle nykytutkimuksen tuloksia. Nykyisen käytössä olevan vanhan kuvailun muuttaminen vastaamaan uusia vaatimuksia ei onnistu todennäköisesti yhtä helposti kuin kirjastokuvailussa tällä hetkellä tehtävä muutostyö. Tämä johtuu ainakin osittain arkistoaineiston luonteesta sekä vanhan ja uuden kuvailutavan suurista eroista.

Kuvailumallin ja -sääntöjen kehittämisen kannalta tarvitaan lisää tutkimusta arkistoaineiston käyttäjien tiedontarpeista ja niihin liittyvistä hakutarpeista. Avoimen linkitetyn datan hyödyntäminen arkistoaineiston kuvailussa ja hakujärjestelmissä vaatii runsaasti selvitystä. Jo nyt erilaisia paikkatieto-ontologioita pyritään hyödyntämään hakujärjestelmissä. Myös muita aiheita käsittelevien ontologioiden hyödyntämisen mahdollisuus täytyy selvittää. Erillinen tapahtuma-ontologia, johon voi linkittää aineistoja tai asiakirjoja hyödyttäisi arkistoaineiston käyttäjiä.

7. JOHTOPÄÄTÖKSET

Tässä luvussa kerrotaan tutkielman tulosten perusteella tehdyistä johtopäätöksistä. Alaluvussa 7.2 on kerrottu tulosten perusteella annettavat suositukset jatkotutkimukselle ja arkistokuvailun kehittämiseksi. Alaluvussa 7.3 käydään läpi tutkielman tuloksiin ja menetelmiin liittyvät rajoitukset.

7.1. Kuvailumallien vaikutuksia aineiston haettavuuteen

Käytännön kuvailutyössä haun kannalta laadukkaamman kuvailun tekeminen helpottuu, kun kuvailtaviin kohteisiin voi liittää yksiselitteisiä arvoja tai attribuutteja. Esimerkiksi valmiiksi kuvaillut toimijat ja aineistoon liittyvä toiminta tai tapahtuma lisää johdonmukaisuutta kuvailuun. Valmiiden arvojen merkitykset ovat ontologioihin tallennettuna lisäksi koneen ymmärtämässä muodossa, mikä osaltaan parantaa hakumahdollisuuksia.

Tämän tutkielman tulosten perusteella nykyisessä arkistokuvailumallissa ei ole otettu riittävästi huomioon aineistojen käyttäjien näkökulmaa aineiston haun kannalta. Arkistoaineistoa haettaessa onkin ennen internet-palveluita tukeuduttu arkistotyöntekijöiden ammattitaitoon aineistojen etsittäessä. AHAA-kuvailumalli tähtää aineistojen parempaan haettavuuteen aineistojen käytön lisäämiseksi. Tässä on tämän tutkielman perusteella onnistuttu hyvin.

AHAA-kuvailumalli määrittää kuvailuelementtien kuvailun erillisinä toisistaan ja niiden yhdistämisen suhteiden avulla. Tällä on monia vaikutuksia aineiston haettavuuteen. Se tuo kuvailuun johdonmukaisuutta ja täsmällisyyttä. Tietojen tallentaminen yksiselitteisesti määriteltynä ja niiden liittäminen aineistoon suhteiden avulla auttaa aineistojen käyttäjiä hakemaan itselleen merkityksellistä aineistoa.

Kuvailutavan vaikutuksista aineiston haettavuuteen käytännössä voidaan arvioida ja tutkia vasta, kun aineistoa on kuvailtu uuden tavan mukaisesti. Suuri rooli on myös tulevilla hakujärjestelmällä ja sen käyttöliittymällä. Myös ontologioiden laatu voi vaikuttaa aineistojen haettavuuteen.

AHAA-kuvailumallin käyttöönotto ei ole vielä varmaa. Myös International Council of Archives -organisaatio on työstänyt uutta kuvailumallia nimeltään Records in Context (RiC), ja tämä työ on otettu huomioon myös Suomessa. Se on suurilta osin samanlainen kuin AHAA-kuvailumalli. (Henttonen & Kilkki, 2015, 3.) Kuvailutapa tulee joka tapauksessa muuttumaan, mutta hakujärjestelmissä ja niiden käyttöliittymissä täytyy ottaa huomioon myös vanhan kuvailutavan mukaisesti kuvaillut aineistot. Uuden kuvailutavan käyttöönottoon tarvitaan runsaasti resursseja ja

se vaikuttaa mm. kuvailijoiden työhön. Olemassa olevien aineistojen kuvailun ontologisointi voi kestää kauan ja nykyiset kuvailutiedot ovat usein riittämättömiä ontologisointiin. Pidemmällä aikavälillä ontologisointi ja muut muutokset tulevat kuitenkin todennäköisesti vähentämään työn määrää mm. vähentämällä päällekkäistä työtä.

Uuden kuvailutavan toteuttamiseksi tarvitaan uusi hakemistojärjestelmä. Finnaan on tarkoitus rakentaa oma näkymä eli käyttöliittymä arkistoaineistojen käyttäjille. Tällä hetkellä arkistoaineistoja voi hakea Finnan kansallisesta näkymästä (finna.fi). Arkistoaineistojen käyttäjien tiedonhakarpeet paremmin huomioiva käyttöliittymä edesauttaa käyttäjien työtä. Käyttäjien tekemällä kuvailulla olisi mahdollista rikastaa vanhojen aineistojen kuvailua ja aineistoihin olisi hyvä pystyä linkittämään tietokannan ulkopuolista informaatiota esimerkiksi internetistä. Ontologioiden käyttöönottoaminen kuvailussa ja hakujärjestelmässä voi kuitenkin olla aluksi hidasta. Osa kuvailusta toteutetaan jo ennen aineiston syntyä ja sen elinkaaren aikana kuvailutieto täydentyy jatkuvasti. Kuvailumetatiedon ontologisointi on tärkeää, vaikka AHAA-kuvailumalli ei tulisiakaan käyttöön. Ontologioilla ja auktorisoiduilla tiedoilla voidaan toteuttaa uuden ajattelutavan mukainen aineiston kontekstualisointi, mikä on kuvailun kannalta olennaisinta muutoksessa. Ajallisten ilmiöiden eli tapahtumien sekä paikkojen (myös historiallisten) kuvaaminen ja niiden esittäminen ontologioiden avulla voisi olla ensimmäinen askel kuvailun rikastamisessa ja kontekstien paremmassa esiintuomisessa.

7.2. Tulosten pohjalta annettavat suositukset

Tulosten perusteella kuvailun tulisi olla mahdollisimman yhtenäistä eri aineistojen kesken. Myös mahdollisimman moni metatieto ja niiden välinen suhde kannattaa asettaa pakollisiksi kuvailussa. Suhteiden luominen aineistojen välille tuo apua erillisten, mutta jollain tavalla yhteen liittyvien aineistojen yhdistämiseen. Arkistoaineiston kuvailussa kontekstin esiin tuominen auttaa tulkitsemaan aineistoa ja osoittaa aineiston autenttisuuden.

Ontologioiden tai muuten semanttisten tietokantojen tuoma etu suhteessa muunlaisiin kontrolloituihin sanastoihin on selkeä. Ontologioissa käsitteiden yksiselitteiset tietokoneenkin ymmärtämät merkitykset ja niiden välinen hierarkia helpottaa sekä kuvailijan työtä että aineistojen käyttäjän tiedonhakua. Linkitetyn (avoimen) datan avulla voidaan yhdistää myös eri arkistoissa sijaitsevia aineistoja toisiinsa esimerkiksi aiheen perusteella. Linkitetyn datan avulla voidaan yhdistää myös muualla internetissä olevaa informaatiota arkistoaineistoihin.

Tulosten perusteella metatietoelementtien välisiä suhteita pitäisi olla monipuolisesti ja lukumäärällisesti paljon. Etenkin eri samaa aihetta käsittelevien tai samaan ajalliseen ilmiöön liittyvien erillisten aineistojen välille pitäisi luoda suhde nimenomaan aihe- -tai tapahtumametatietojen kautta. Aihetta käsitteleviä metatietoja voi olla useita erilaisia ja ne voivat olla kontrolloituja metatietoja, jolloin suhteiden luominen helpottuu. Historioitsijoiden ja muiden tutkijoiden kannalta tällä on suuri merkitys, sillä he pyrkivät usein luomaan kokonaiskuvan tutkimastaan aiheesta. Tällöin kaikki aiheeseen liittyvä aineisto voi olla merkityksellistä tutkimuksen kannalta.

Käyttäjien tekemästä passiivisesta (automaattinen) tai aktiivisesta kuvailusta olisi hyötyä kuvailun rikastamisessa. Aineistojen ulkopuolisten tapahtumien ja ajallisten ilmiöiden liittäminen aineistoon voisi olla aineistojen käyttäjien aktiivisesti tuottamaa kuvailutietoa. Tulevan Finna-arkistokäyttöliittymän ja sen hakujärjestelmän olisi hyvä ottaa huomioon tämä.

7.3 Tuloksiin ja menetelmiin liittyvät rajoitukset

Arkistoaineiston käyttäjiin kohdistuvaa tutkimusta on tehty vain rajallinen määrä, joten tähän tutkielmaan löydetty hakukriteerit voisivat olla erilaiset. Tähän tutkielmaan valikoituivat mukaan ne tiedonhaketutkimukset, joihin kirjoittajalla oli pääsy sähköisesti.

Hakukriteerien vertailu kuvailumetatietoihin ja kuvailutapoihin on osittain tulkinnallista. Tätä tulkinnanvaraisuutta on pyritty kuitenkin avaamaan jo tulosluvussa. Arkistoaineiston haettavuuteen ja löydettävyyteen vaikuttaa moni eri asia. Tässä tutkielmassa kuvailumallien sekä niistä johdettujen kuvailusääntöjen ja hakemiskriteerien välistä jätettiin pois hakujärjestelmä ja sen käyttöliittymä sekä tiedon varsinainen tallennustapa. Niillä voidaan osaltaan vaikuttaa suuresti aineiston haun onnistumiseen.

Nykyiset kuvailusäännöt ja keskeneräiset AHAA-kuvailusäännöt ovat tässä tutkielmassa hieman eri asemassa, koska AHAA-kuvailusäännöt eivät ota kantaa metatietojen pakollisuuteen tätä kirjoitettaessa.

LÄHTEET

- Adams, M. O. (2007). Analyzing archives and finding facts: use and users of digital data records. *Archival Science*, 7(1), 21–36. doi:10.1007/s10502-007-9056-4
- Alaterä, A., & Halttunen, K. (2002). *Tiedonhaun perusteet - osa lukutaitoa*. Helsinki: BTJ Kirjastopalvelu Oy.
- Arkistolaitos. (1997a). *Arkistojen kuvailu- ja luettelointisäännöt*. Helsinki: Arkistolaitos.
- Arkistolaitos. (1997b). *Arkistojen kuvailu- ja luettelointisäännöt 1997*. Helsinki: Arkistolaitos. Retrieved November 15, 2014, from <http://www.arkisto.fi/fi/palvelut/normit/suositukset/arkistojen-kuvailu--ja-luettelointisaaennot/>
- Arkistolaitos. (2006). 5.3.4 Kuvailu. *Arkistot - yhteiskunnan toimiva muisti*. Retrieved December 12, 2014, from <http://www.arkisto.fi/fi/palvelut/julkaisuluettelo/d-verkko-oppaat/arkistot-yhteiskunnan-toimiva-muisti/arkistojen-hallinta-ja-kaeytettaevyyden-parantaminen-2>
- Arkistolaitos. (2009a). *Sähke_Liite2_Metatietomalli*. Helsinki: Arkistolaitos.
- Arkistolaitos. (2009b). *Sähköisten asiakirjallisten tietojen käsittely , hallinta ja säilyttäminen*. Helsinki: Arkistolaitos.
- Arkistolaitos. (2013a). *Arkistowiki: Proveniensi*. Retrieved March 1, 2015, from <http://wiki.narc.fi/arkistowiki/index.php/Proveniensi>
- Arkistolaitos. (2013b). *Sähköisen tiedon ohjaus*. Retrieved February 1, 2015, from <http://www.arkisto.fi/fi/palvelut/julkisen-hallinnon-saehkoeiset-palvelut/saehkoeisen-tiedon-ohjaus/>
- Arkistolaitos. (2014). *Arkistoille yhteinen hakemistopalvelu AHAA*. Retrieved February 1, 2015, from <http://www.arkisto.fi/fi/palvelut/arkistoyhteistyoe/arkistojen-hakemistopalvelu>
- Bak, G. (2012). Continuous classification: capturing dynamic relationships among information resources. *Archival Science*, 12(3), 287–318. doi:10.1007/s10502-012-9171-8
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. doi:doi:10.4018/jswis.2009081901
- Bunn, J. (2013). Developing descriptive standards: a renewed call to action. *Archives and Records: The Journal of the Archives and Records Association*, 34(2), 235–247. doi:10.1080/23257962.2013.830066
- Coyle, K. (2010). FRBR , the Domain Model. *Library Technology Reports*, 46(2), 20–25.
- Duff, W. M., & Johnson, C. A. (2003). Where Is the List with All the Names? Information-Seeking Behavior of Genealogists. *The American Archivist*, 66, 79–95.

- Elena, T., Katifori, A., Vassilakis, C., Lepouras, G., & Halatsis, C. (2010). Historical research in archives: User methodology and supporting tools. *International Journal on Digital Libraries*, 11(2010), 25–36. doi:10.1007/s00799-010-0062-4
- Finna. (2013). Retrieved December 1, 2014, from <https://www.finna.fi/>
- Georgas, H. (2013). Google vs. the Library: Student Preferences and Perceptions When Doing Research Using Google and a Federated Search Tool. *Libraries and the Academy*, 13(2), 165–185. doi:10.1353/pla.2013.0011
- Georgas, H. (2014). Google vs. the Library (Part II): Student Search Patterns and Behaviors when Using Google and a Federated Search Tool. *Libraries and the Academy*, 14(4), 503–532. doi:10.1353/pla.2014.0034
- Gracy, K. F. (2014). Archival description and linked data: a preliminary study of opportunities and implementation challenges. *Archival Science*. doi:10.1007/s10502-014-9216-2
- Gueguen, G., Manoel, V., Pitti, D. V., & Grimoüard, C. S. (2013). Toward an International Conceptual Model for Archival Description : A Preliminary Report from the International Council on Archives ' Experts Group on Archival Description. *The American Archivist*, 76(2), 567–584.
- Haasio, A., & Savolainen, R. (2004). *Tiedonhankintatutkimuksen perusteet*. Helsinki: BTJ Kirjastopalvelu Oy.
- Hamburger, S. (2004). How Researchers Search for Manuscript and Archival Collections. *Journal of Archival Organization*, 2, 79–102. doi:10.1300/J201v02n01_07
- Henttonen, P. (2012). Arkistojen kuvailu- ja luettelointi- sääntöjen kehittäminen - kansallinen ponnistus vai kansainvälistä yhteistyötä ?. *Informaatiotutkimus*. 31(2), 1–8.
- Henttonen, P., & Kilkki, J. (2013). *Arkistokuvailun kansallinen käsitelmä 0.15 (luonnos)*.
- Henttonen, P., & Kilkki, J. (2015). *Arkistokuvailun kansallinen käsitelmä luonnosversio 0.2*. Helsinki: Arkistolaitos.
- Hill, A. (2004). Serving the invisible researcher: Meeting the needs of online users. *Journal of the Society of Archivists*, 25(2), 139–148. doi:10.1080/0037981042000271466
- Hormia-Poutanen, K. (2013). *Finna – tietoa tarvitseville ja elämyksiä etsiville*. Kansalliskirjasto, 1. Retrieved from http://www.kansalliskirjasto.fi/attachments/5xaqbsRw7/6FwXzhj2j/Files/CurrentFile/Kansalliskirjasto_1_2013.pdf
- Huvila, I. (2008). Participatory archive: towards decentralised curation, radical user orientation, and broader contextualisation of records management. *Archival Science*, 8(1), 15–36. doi:10.1007/s10502-008-9071-0

- Jain, V., & Singh, M. (2013). Ontology Based Information Retrieval in Semantic Web: A Survey. *International Journal of Information Technology and Computer Science*, 5(September), 62–69. doi:10.5815/ijitcs.2013.10.06
- JHS 143. (2006). Asiakirjojen kuvailun ja hallinnan metatiedot. Helsinki: Arkistolaitos.
- Kamut. (1997). Kamut-tietorakenne: kirjastojen arkistojen ja taide- sekä kulttuurihistoriallisten museoiden yhteiskäyttöiset luettelointitiedot. Projektin loppuraportti 19.3.1997.
- Kettunen, K., & Henttonen, P. (2010). Missing in action? Content of records management metadata in real life. *Library & Information Science Research*, 32(1), 43–52. doi:10.1016/j.lisr.2009.10.002
- Kuvailusäännöt-Tietovaranto - Esimerkit tänne. (2013).
- Lampela, A. (2013). Kuvailusäännöt_Sisältö- ja rakennetietojen alue.
- Lindquist, T., Dulock, M., Törnroos, J., Hyvönen, E., & Mäkelä, E. (2013). Using Linked Open Data to Enhance Subject Access in Online Primary Sources. *Cataloging & Classification Quarterly*, 51(8), 913–928. doi:10.1080/01639374.2013.823583
- Lindquist, T., & Long, H. (2011). How can educational technology facilitate student engagement with online primary sources? *Library Hi Tech*, 29(2), 224–241. doi:10.1108/07378831111138152
- M. Keenan, T. (2014). Resource description and access: cataloging standards affect reference service. *Reference Services Review*, 42(3), 446–466. doi:10.1108/RSR-04-2014-0008
- Nimer, C. (2010). RDA and Archives. *Journal of Archival Organization*, 8(3), 227–243. doi:10.1080/15332748.2010.550799
- Nykyri, S., & Palonen, T. (2014). Ontologioiden käytön moninaisuudesta ja tulevaisuusnäköymistä : Finto-palvelun sidosryhmien ontologiapalvelulle kohdistamat kehitystoiveet. Helsinki.
- Parandjuk, J. C. (2010). Using Information Architecture to Evaluate Digital Libraries. *The Reference Librarian*, 51(2), 124–134. doi:10.1080/02763870903579737
- Schaffner, J. (2009). The Metadata is the Interface of Archives and Special Collections , Synthesized from User Studies. OCLC Research.
- Sinn, D., & Soares, N. (2014). Historians ' Use of Digital Archival Collections : The Web , Historical Scholarship , and Archival Research. *Journal of the Association for Information Science and Technology*, 1–16. doi:10.1002/asi
- Suominen, V., Saarti, J., & Tuomi, P. (2009). Bibliografisen valvonta Johdatus luetteloinnin ja sisällönkuvailun menetelmiin. Helsinki: BTJ Finland Oy.
- Thurman, A. C. (2007). FRBR and archival materials. In A. G. Taylor (Ed.), *Understanding FRBR: what it is and how it will affect our retrieval tools* (pp. 97–102). Westport, Conn.: Libraries Unlimited.

Toiminta30052013 (2013).

Trombone, A. (2014). New display models of bibliographic data and resources : cataloguing / resource description and search results communication, 5(2), 19–31. doi:10.4403/jlis.it-10063

Vakkari, M., Simovaara, & Valanto, S. (2004). KAMUT 2. Muistiorganisaatioiden tietovarannot yhteiskäyttöön. Toteuttamismahdollisuudet ja toimenpidesuositus. Museovirasto.

Van Ballegooie, M., & Borie, J. (2014). From Record-Bound to Boundless: FRBR, Linked Data, and New Possibilities for Serials Cataloging. *The Serials Librarian*, 66(1-4), 76–87. doi:10.1080/0361526X.2014.879527